Sparse Nonstationary Gabor Expansions - with Applications to Music Signals

Ottosen, Emil Solsbæk

DOI (link to publication from Publisher):
10.5278/VBN.PHD.ENG.00041

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

SPARSE NONSTATIONARY GABOR EXPANSIONS

WITH APPLICATIONS TO MUSIC SIGNALS

BY
EMIL SOLSBÆK OTTOSEN

DISSERTATION SUBMITTED 2018

AALBORG UNIVERSITY
DENMARK

Sparse Nonstationary Gabor Expansions

with Applications to Music Signals

Ph.D. Dissertation

Emil Solsbæk Ottosen

Dissertation submitted February 12, 2018

Dissertation submitted: February 12, 2018

PhD supervisor: Prof. Morten Nielsen
Dept. of Mathematical Sciences
Aalborg University

PhD committee: Professor Horia Cornean (Chairman)
Aalborg University

Associate Professor Jakob Lemvig
Technical University of Denmark

Professor Hans G. Feichtinger
University of Vienna

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Mathematical Sciences

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-153-8

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2018

Preface

The work presented in this thesis was carried out at the Department of Mathematical Sciences, Aalborg University, within the period February 2015 to February 2018. The Ph.D. project has been supervised by professor Morten Nielsen from the Department of Mathematical Sciences, Aalborg University. Parts of the research was accomplished during a research stay at the Faculty of Mathematics, University of Vienna, within the period September 2016 to December 2016. During the research stay, the Ph.D. project was supervised by Dr. Monika Dörfler from the research group NuHAG (Numeric Harmonic Analysis Group).

The central part of this thesis consists of a collection of four scientific papers. Three of these papers have been published in peer reviewed journals while the remaining one is undergoing review at the time of submission. The first part of the thesis gives an introduction to time-frequency analysis and provides some background material necessary for understanding the problems considered in the scientific papers. The second part of the thesis presents the four scientific papers as self-contained articles with separate abstracts, introductions, bibliographies and so on.

First of all, I would like to thank my supervisor Morten Nielsen for introducing me to the field of time-frequency analysis and for his guidance throughout the Ph.D. project. I would also like to thank Monika Dörfler, and the other members of NuHAG, for an excellent and constructive stay at the University of Vienna. My colleges at the Department of Mathematical Sciences also deserves recognition for their help, especially Søren Vilsen for assistance on various computer problems. Finally, I would like to thank my friends and family for their support during the Ph.D. project.

Emil Solsbæk Ottosen
Aalborg University, February 12, 2018

Thesis Details

Thesis title:	Sparse Nonstationary Gabor Expansions — with Applications to Music Signals
PhD student:	Emil Solsbæk Ottosen Dept. of Mathematical Sciences Aalborg University
PhD Supervisor:	Prof. Morten Nielsen Dept. of Mathematical Sciences Aalborg University

The thesis is based on the following four scientific papers:

(A) E. S. Ottosen and M. Nielsen. Weighted Thresholding and Nonlinear Approximation. Submitted to *Indian Journal of Pure and Applied Mathematics*, 2017.

(B) E. S. Ottosen and M. Nielsen. A Characterization of Sparse Nonstationary Gabor Expansions. Accepted for publication in *Journal of Fourier Analysis and Applications*, 2017. Available: `https://link.springer.com/article/10.1007%2Fs00041-017-9546-6`.

(C) E. S. Ottosen and M. Nielsen. Nonlinear Approximation with Nonstationary Gabor Frames. Accepted for publication in *Advances in Computational Mathematics*, 2017. Available: `https://doi.org/10.1007/s10444-017-9577-1`

(D) E. S. Ottosen and M. Dörfler. A Phase Vocoder Based on Nonstationary Gabor Frames. Published in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol 25, no. 11, pp. 2199–2208, 2017. Available: `http://ieeexplore.ieee.org/document/8031036/`.

English Summary

In this thesis we consider sparseness properties of classical Gabor expansions and adaptive nonstationary Gabor expansions. A classical Gabor expansion decomposes a signal into a convergent sum of time-frequency (TF) localized atoms, which are constructed as TF-shifts of a single fixed window function. In contrast, nonstationary Gabor expansions allow for the usage of multiple window functions to obtain an adaptive TF resolution. Both classical and nonstationary Gabor expansions have proven to be extremely useful for representing music signals as they provide sparse and structured TF representations. Sparseness of a TF representation is desirable since it reduces the computational cost involved in processing the expansion coefficients. Also, the sparseness property allows for efficient approximations of the signal by thresholding the associated expansion coefficients. This kind of approximation belongs to the area of approximation theory known as nonlinear approximation.

In paper A of this thesis we consider sparseness properties of classical Gabor expansions in the general framework of nonlinear approximation theory. The main contribution of this paper is an upper bound (a Jackson inequality) on the approximation error occurring when thresholding the expansion coefficients with respect to certain weight functions. These weight functions seek to exploit the coherence between expansion coefficients which is not accounted for in traditional greedy algorithms.

Nonstationary Gabor frames extend the concept of classical Gabor frames by allowing for a flexible TF resolution along either the time or the frequency axis. In paper B of this thesis we use decomposition spaces to characterize those signals which permit sparse expansions with respect to certain nonstationary Gabor frames with flexible frequency resolution. In paper C we consider a similar approach for nonstationary Gabor frames with flexible time resolution and provide a numerical analysis of the associated approximation rate. Finally, in paper D we consider a practical application and construct a new time-stretching algorithm based on the theory of nonstationary Gabor frames. Time-stretching is the task of modifying the length of a signal without affecting its frequencies.

Dansk Resumé

I denne afhandling undersøger vi sparseness egenskaber ved klassiske Gabor udviklinger og adaptive ikke-stationære Gabor udviklinger. En klassisk Gabor udvikling nedbryder et signal til en konvergent sum af tids-frekvens (TF) lokaliserede atomer, der er konstruerede som TF-skift af en enkelt vinduesfunktion. I modsætning hertil så anvender ikke-stationære Gabor udviklinger flere vinduesfunktioner til at opnå en adaptiv TF opløsning. Både klassiske og ikke-stationære Gabor udviklinger har vist sig at være særligt nyttige til at repræsentere musiksignaler, da de producerer sparse og strukturerede TF repræsentationer. Sparseness af en TF repræsentation er ønskværdigt, da det nedsætter mængden af computerkraft, der skal bruges til at bearbejde koefficienterne. Desuden medfører sparseness egenskaben, at man effektivt kan approksimere signalet ved at thresholde de tilhørende koefficienter. Denne type approksimation tilhører det felt af approksimationsteorien, der kendes som ikke-linear approksimation.

I artikel A af denne afhandling undersøger vi sparseness egenskaber ved klassiske Gabor udviklinger i det generelle framework, som benyttes i ikkelineær approksimationsteory. Hovedresultatet af denne artikel er en øvre grænse (en Jackson ulighed) på den approximationsfejl, der opstår, når man thresholder koefficienterne mht. visse vægtfunktioner. Disse vægtfunktioner forsøger at udnytte sammenhængen mellem koefficienterne, hvilket der almindeligvis ikke tages højde for i traditionelle grådige algoritmer.

Ikke-stationære Gabor frames generaliserer klassiske Gabor frames ved at tillade en fleksibel TF opløsning langs enten tids- eller frekvensaksen. I artikel B af denne afhandling bruger vi dekompositionsrum til at karakterisere de signaler, der har sparse udviklinger mht. ikke-stationære Gabor frames med fleksibel opløsning i frekvens. I artikel C benytter vi en lignende tilgang for ikke-stationære Gabor frames med fleksibel opløsning i tid og inkluderer desuden en numeriske analyse af den tilhørende approksimationsrate. Til sidst, i artikel D, ser vi på en praktisk anvendelse og konstruerer en ny time-stretching algoritme baseret på teorien om ikke-stationære Gabor frames. Time-stretching er anvendelsen, hvor man modificerer længden af et signal uden at ændre ved dets frekvenser.

Contents

Part I

Introduction

Introduction

This thesis investigates sparseness properties of time-frequency (TF) representations obtain with classical Gabor frames and adaptive nonstationary Gabor frames. In particular, the theory is relevant for analyzing music signals as these signals tend to produce sparse TF representations when expanded in a (nonstationary) Gabor dictionary. In this first part of the thesis we give an introduction to TF analysis in the general framework provided by frame theory. Both classical Gabor frames and nonstationary Gabor frames are (as the words suggest) special cases of frames and it therefore make sense to consider them in a unified framework. To illustrate the concepts from TF analysis we include a real world example and analyze the associated TF representations. With this approach, the abstract theory presented in the papers in the second part of the thesis will hopefully be easier to grasp as many of the deep theoretical concepts take their roots in practical problems from TF analysis.

1 Frame theory

The main property of a frame $\{g_k\}_{k\in\mathbb{N}}$ in a separable infinite dimensional Hilbert space $\mathcal{H}$ is that every $f \in \mathcal{H}$ has a stable expansion of the form $f = \sum_{k\in\mathbb{N}} c_k g_k$ with $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$. In general, frames are redundant systems with non-unique expansion coefficients $\{c_k\}_{k\in\mathbb{N}}$. The notion of frames was first introduced in 1952 by Duffin and Schaeffer [35] who studied the properties of overcomplete families of exponential functions. Much later in 1986, Daubechies, Grossmann, and Meyer [19] revisited the theory and considered new types of frames not restricted to the framework of nonharmonic Fourier series. We also mention the work by Young [109], Heil and Walnut [63], and Daubechies [16, 17] as examples of some of the early work done on frame theory. A modern and thorough introduction to frame theory can be found in the book by Christensen [13].

Formally, a sequence $\{g_k\}_{k\in\mathbb{N}} \subset \mathcal{H}$ is a frame for $\mathcal{H}$ if there exist frame bounds $A, B > 0$ such that

$$A \|f\|_{\mathcal{H}}^2 \leq \sum_{k\in\mathbb{N}} |\langle f, g_k\rangle|^2 \leq B \|f\|_{\mathcal{H}}^2, \quad \forall f \in \mathcal{H}.$$

If $A = B$ then the frame is said to be tight. It follows from Parseval's identity that any orthonormal basis is a tight frame with $A = B = 1$. For a given frame $\{g_k\}_{k\in\mathbb{N}} \subset \mathcal{H}$, we define the associated frame operator $S : \mathcal{H} \to \mathcal{H}$ by

$$Sf = \sum_{k\in\mathbb{N}} \langle f, g_k\rangle g_k, \quad f \in \mathcal{H}.$$

The frame property implies that S is a bounded, invertible, self-adjoint, and positive operator, which permits a dual frame $\{\tilde{g}_k\}_{k\in\mathbb{N}} := \{S^{-1}g_k\}_{k\in\mathbb{N}}$ which is also a frame but with frame-bounds $B^{-1}, A^{-1} > 0$ [13, 55, 61, 86]. It follows that every $f \in \mathcal{H}$ has a frame expansion of the form

$$f = SS^{-1}f = \sum_{k\in\mathbb{N}} \left\langle S^{-1}f, g_k\right\rangle g_k = \sum_{k\in\mathbb{N}} \langle f, \tilde{g}_k\rangle g_k,$$

with unconditional convergence in $\mathcal{H}$. Similarly, f possesses an expansion with respect to the dual frame $\{\tilde{g}_k\}_{k\in\mathbb{N}}$ in the sense that

$$f = S^{-1}Sf = \sum_{k\in\mathbb{N}} \langle f, g_k\rangle S^{-1}g_k = \sum_{k\in\mathbb{N}} \langle f, g_k\rangle \tilde{g}_k. \tag{1}$$

We choose to work mainly with the expansion given in (1). The frame coefficients $\{\langle f, g_k\rangle\}_{k\in\mathbb{N}}$ can be shown to minimize the ℓ^2-norm among all possible reconstruction coefficients $\{c_k\}_{k\in\mathbb{N}}$ satisfying $f = \sum_{k\in\mathbb{N}} c_k\tilde{g}_k$, see [109].

For a tight frame we note that $\langle Sf, f\rangle = \sum_{k\in\mathbb{N}} |\langle f, g_k\rangle|^2 = A\|f\|^2$, which implies $S = AI$ with I denoting the identity operator on $\mathcal{H}$. Hence, for an orthonormal basis we obtain $S = I$ and the frame expansion in (1) reduces to the well-known unique reconstruction formula $f = \sum_{k\in\mathbb{N}}\langle f, g_k\rangle g_k$. However, for many practical purposes it is beneficial to consider redundant frames where the reconstruction coefficients are not uniquely determined [14, 27, 95].

2 Gabor frames

In this section we consider an important kind of frames for $L^2(\mathbb{R}^d)$ known as Gabor frames. These frames are named after D. Gabor [52] who in 1946 considered a new approach for signal decomposition using TF localized atoms. These decompositions was later studied by Janssen [70, 71] in the early 1980s and was combined with frame theory by Daubechies, Grossmann, and Meyer [19] in their paper from 1986. Further studies on the fundamental properties of Gabor frames were performed in the late 1980s by Feichtinger and

Gröchenig [40, 42, 45, 46], in the early 1990s by Daubechies [16] and Heil and Walnut [63, 106], and later in the 1990s by several other authors (see [20, 73, 96] and references therein). We also mention the two books [47, 48] by Feichtinger and Strohmer which contain surveys on various topics of Gabor analysis.

Gabor frames are based on two classes of unitary operators on $L^2(\mathbb{R}^d)$, namely translation and modulation. For $\alpha, \beta \in \mathbb{R}^d$ we define the translation operator $T_\alpha : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ and the modulation operator $M_\beta : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ by

$$T_\alpha f(x) = f(x-\alpha) \quad \text{and} \quad M_\beta f(x) = f(x)e^{2\pi i \beta \cdot x}.$$

Given lattice parameters $a, b > 0$, and a fixed window function $g \in L^2(\mathbb{R}^d)$, we define the Gabor system $\{g_{m,n}\}_{m,n\in\mathbb{Z}^d} := \{M_{mb}T_{na}g\}_{m,n\in\mathbb{Z}^d}$. If the system constitutes a frame for $L^2(\mathbb{R}^d)$ then it is referred to as a Gabor frame. Explicitly written, the frame elements of a Gabor frame are given by

$$g_{m,n}(x) = g(x-na)e^{2\pi i mb\cdot x}, \quad m,n \in \mathbb{Z}^d,$$

and the associated frame operator by

$$Sf = \sum_{m,n\in\mathbb{Z}^d} \langle f, g_{m,n}\rangle\, g_{m,n}, \quad f \in L^2(\mathbb{R}^d).$$

By cumbersome calculations it can be shown that S (and consequently S^{-1}) commutes with translation and modulation [55]. The dual frame of $\{g_{m,n}\}_{m,n}$ is therefore given by

$$\{\tilde{g}_{m,n}\}_{m,n} = \{S^{-1}g_{m,n}\}_{m,n} = \{M_{mb}T_{na}S^{-1}g\}_{m,n} = \{M_{mb}T_{na}\tilde{g}\}_{m,n},$$

with $\tilde{g} := S^{-1}g$ denoting the so-called dual window of g. The fact that the dual frame is also a Gabor frame is a special property of Gabor frames not shared by general frames. The frame expansions in (1) take the form

$$f = \sum_{m,n\in\mathbb{Z}^d} \langle f, g_{m,n}\rangle\, \tilde{g}_{m,n}, \quad f \in L^2(\mathbb{R}^d).$$

The Gabor frames we consider here are often referred to as regular Gabor frames since the sampling points $\{na, mb\}_{m,n\in\mathbb{Z}^d}$ form a separable lattice $a\mathbb{Z}^d \times b\mathbb{Z}^d$ in $\mathbb{R}^{2d}$. The lattice parameters $a, b > 0$ determine the density of the Gabor frame, which is usually divided into three cases:

1. $ab > 1$: This is called undersampling and implies that $\{g_{m,n}\}_{m,n}$ cannot form a frame for $L^2(\mathbb{R}^d)$. This was proved for the rational case $ab \in \mathbb{Q}$ by Daubechies [16] and for the general case by Baggett [5] (see also the paper by Janssen [72]).

2. $ab = 1$: This is called critical sampling. In this case, $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ if and only if it is also a Riesz basis, i.e., a basis of the form $\{Ue_k\}_{k\in\mathbb{N}}$ with $\{e_k\}_{k\in\mathbb{N}}$ being an orthonormal basis and U a bounded bijective operator on $L^2(\mathbb{R}^d)$, see [13]. Additionally, if $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ then the Balian-Low theorem states that g cannot be well localized in both time and frequency [7, 9, 84]. We elaborate further on this point in Section 4.1.

3. $ab < 1$: This is called oversampling. In this case, if $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ then it cannot be a Riesz basis. A frame which is not a Riesz basis is called redundant since there exist coefficients $\{c_{m,n}\}_{m,n}$ in $\ell^2 \setminus \{0\}$ with $\sum_{m,n} c_{m,n}g_{m,n} = 0$, see [13].

The original expansion considered by D. Gabor in 1946 was in the critical case with $ab = 1$ and $g(x) = e^{-x^2/2}$ being the Gaussian function [52]. This expansion has later been shown to be unstable as the Gaussian window posses the special property that the corresponding Gabor system $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ if and only if $ab < 1$. This result was first conjectured by Daubechies and Grossmann [18] in 1988 and later proven in 1992 independently by Lyubarskiĭ [85] and Seip and Wallstén [99, 100]. For a general window function $g \in L^2(\mathbb{R}^d)$ it is difficult to find an exact range of parameters $a, b > 0$ guaranteeing that $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ [56, 75, 81]. It is therefore important to note that the terminology introduced above is slightly misleading. Even with heavy oversampling $ab < 1$ we are not guaranteed that $\{g_{m,n}\}_{m,n}$ forms a frame for $L^2(\mathbb{R}^d)$.

We now consider a simple construction of Gabor frames, which is of great practical importance. Let $a, b > 0$ and assume $g \in L^2(\mathbb{R}^d)$ satisfies $\text{supp}(g) \subseteq [0, 1/b]^d$. With $G(x) := b^{-d} \sum_{n\in\mathbb{Z}^d} |g(x - na)|^2$, the frame operator for $\{g_{m,n}\}_{m,n}$ turns out to be the multiplication operator [19]

$$Sf(x) = G(x)f(x), \quad f \in L^2(\mathbb{R}^d).$$

Consequently, $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ with frame bounds $A, B > 0$ if and only if

$$A \leq G(x) \leq B, \quad \text{for a.e. } x \in \mathbb{R}^d.$$

Furthermore, if $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ then the dual frame is given by $\tilde{g}_{m,n}(x) = G^{-1}(x)g_{m,n}(x)$. Such frames are traditionally referred to as painless nonorthogonal expansions [19]. To be consistent with the notation in Section 3 we choose to simply call them painless Gabor frames.

3 Nonstationary Gabor frames

One of the shortcomings of classical Gabor frames is the stationarity resulting from applying only one window function. A straightforward generalization of the theory is to choose a countable set of window functions $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$, with corresponding sampling parameters $\{b_n\}_{n\in\mathbb{Z}^d} \subset \mathbb{R}_+$, and then define the system $\{g_{m,n}\}_{m,n\in\mathbb{Z}^d}$ by

$$g_{m,n}(x) = M_{mb_n} g_n(x) = g_n(x) e^{2\pi i m b_n \cdot x}, \quad m, n \in \mathbb{Z}^d. \tag{2}$$

We obtain classical Gabor systems by choosing, for all $n \in \mathbb{Z}^d$, $g_n := T_{na} g$ and $b_n := b$ with $g \in L^2(\mathbb{R}^d)$ and $a, b > 0$.

The generalized Gabor systems defined in (2) were originally studied by Hernández, Labate, and Weiss [64] and later by Ron and Shen [97] under the name of generalized-shift invariant systems (see also [13, 69, 74, 80]). Technically, generalized-shift invariant systems are obtained by applying a Fourier transform to the systems in (2), thereby producing the equivalent systems defined in Section 3.2. The systems in (2) are defined in the time domain whereas the systems in Section 3.2 are defined in the frequency domain. The term nonstationary Gabor frames (NSGFs) was introduced by Jaillet [68] for frames of the form (2) and further used by Balazs et al. [6], Holighaus [65], and Dörfler and Matusiak [33, 34]. We choose to work with the terminology of [68] to emphasize the connection with classical Gabor frames. Likewise, we will often refer to a classical Gabor frame as a stationary Gabor frame.

The important paper [6] by Balazs, Dörfler, Jaillet, Holighaus and Velasco contains the first practical implementations of NSGFs. This paper is accompanied by a Matlab toolbox, which provides source code for construction NSGFs in both the time domain and the frequency domain. The main author of this toolbox is Holighaus and the source code applies routines from the Large Time-Frequency Analysis Toolbox (LTFAT), which was originally founded by Søndergaard [101] and is currently maintained by Průša [94]. The practical implementations presented in [6] shows that NSGFs can be used to create fast adaptive TF representations, which for certain signal classes outperforms classical (stationary) Gabor frames.

Finally, it should be noted that NSGFs can be considered a special case of the more general concept of multi-window Gabor frames [30, 108, 110]. However, so far NSGFs are the only realization of multi-window Gabor frames which permit perfect reconstruction and a fast implementation based on the FFT [83]. These properties makes NSGFs particular interesting from a practical point of view.

3.1 NSGFs in the time domain

We first describe NSGFs in the time domain, i.e., NSGFs of the form (2). The frame operator for a NSGF $\{g_{m,n}\}_{m,n}$ is defined by

$$Sf = \sum_{m,n\in\mathbb{Z}^d} \langle f, g_{m,n}\rangle\, g_{m,n}, \quad f \in L^2(\mathbb{R}^d),$$

and the resulting frame expansions take the form

$$f = \sum_{m,n\in\mathbb{Z}^d} \langle f, g_{m,n}\rangle\, S^{-1}g_{m,n}, \quad f \in L^2(\mathbb{R}^d).$$

In general, the dual frame $\{S^{-1}g_{m,n}\}_{m,n}$ does not produce a new NSGF as is the case for stationary Gabor frames [65]. However, by generalizing the painless condition to the nonstationary case we do obtain dual frames with this property. Let $\{b_n\}_{n\in\mathbb{Z}^d} \subset \mathbb{R}_+$ and assume $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ satisfy $\mathrm{supp}(g_n) \subseteq [0, 1/b_n]^d + a_n$ with $a_n \in \mathbb{R}^d$ for all $n \in \mathbb{Z}^d$. With $G(x) := \sum_{n\in\mathbb{Z}^d} b_n^{-d}|g_n(x)|^2$, the frame operator for $\{g_{m,n}\}_{m,n}$ is the multiplication operator [6, 19]

$$Sf(x) = G(x)f(x), \quad f \in L^2(\mathbb{R}^d).$$

It follows that $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ with frame bounds $A, B > 0$ if and only if

$$A \le G(x) \le B, \quad \text{for a.e. } x \in \mathbb{R}^d.$$

If $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ then the dual frame is given by

$$\tilde{g}_{m,n}(x) = M_{mb_n}\frac{g_n(x)}{G(x)}, \quad x \in \mathbb{R}^d.$$

We note that this result completely generalizes the painless case for stationary Gabor frames. For this reason we refer to the system $\{g_{m,n}\}_{m,n}$ as a painless NSGF [6].

3.2 NSGFs in the frequency domain

As mentioned in the beginning of Section 3, NSGFs have an equivalent implementation in the frequency domain. We use the following normalization for the Fourier transform

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-2\pi i x\cdot\xi}dx, \quad f \in L^1(\mathbb{R}^d),$$

which by standard arguments extends to a unitary operator on $L^2(\mathbb{R}^d)$, see [82]. Given $\{h_m\}_{m\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ and $\{a_m\}_{m\in\mathbb{Z}^d} \subset \mathbb{R}_+$, we define the system $\{h_{m,n}\}_{m,n\in\mathbb{Z}^d}$ by

$$h_{m,n}(x) = T_{na_m}h_m(x) = h_m(x - na_m), \quad m, n \in \mathbb{Z}^d.$$

In complete analogy with the painless condition in the time domain, we define $H(\xi) := \sum_{m\in\mathbb{Z}^d} a_m^{-d}|\hat{h}_m(\xi)|^2$ and assume $\mathrm{supp}(\hat{h}_m) \subseteq [0, 1/a_m]^d + b_m$ with $b_m \in \mathbb{R}^d$ for all $m \in \mathbb{Z}^d$. The frame operator is then given by [6, 19]

$$Sf(x) = (\mathcal{F}^{-1}H * f)(x), \quad f \in L^2(\mathbb{R}^d),$$

which corresponds to a multiplication operator in the frequency domain. Consequently, $\{h_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ with frame bounds $A, B > 0$ if and only if

$$A \leq H(\xi) \leq B, \quad \text{for a.e. } \xi \in \mathbb{R}^d.$$

If $\{h_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ then the dual frame is given by

$$\tilde{h}_{m,n}(x) = T_{na_m} \mathcal{F}^{-1} \left(H^{-1} \hat{h}_m \right)(x), \quad x \in \mathbb{R}^d.$$

In the next section we consider applications of stationary Gabor frames and NSGFs in connection with TF analysis.

4 Time-frequency analysis

The purpose of TF analysis is to combine the information of a signal $f \in L^2(\mathbb{R}^d)$ and its Fourier transform $\hat{f} \in L^2(\mathbb{R}^d)$ into a $2d-$dimensional TF representation $V(x, \xi)$ containing information about the frequencies ξ occurring at time x. A common analogy for a TF representation is the musical score [21]. In Fig. 1 we see the musical score for the first 4 bars of a piece of piano music.

Fig. 1: First 4 bars of "Moanin" by Bobby Timmons.

A musician reads the musical score from left to right and the vertical position of each note determines the pitch of the note. A note on a piano corresponds to a fundamental frequency and overtones with frequencies that are integer multiples of the fundamental frequency. It is the presence of overtones that determines the particular timbre of the instrument. The first note of the musical score in Fig. 1 is F4 which has a fundamental frequency of 350 Hz and overtones of frequencies 700 Hz, 1050 Hz, 1400 Hz, and so on. Suppose the piano music has been recorded with a microphone. This produces a signal $f(x)$ describing the changes over time in air pressure as shown in Fig. 2.

We might be able to determine some rhythmical patterns or maybe even the tempo (♩ $= 126$) by analyzing $f(x)$ directly but we cannot say anything about the melody. On the other hand, calculating the (normalized) values of $|\hat{f}(\xi)|$ we obtain the plot in Fig. 3.

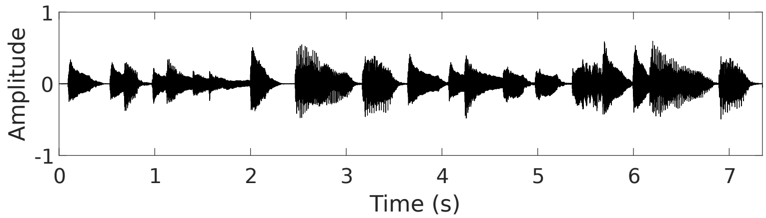

Fig. 2: Time-domain plot of the piano signal.

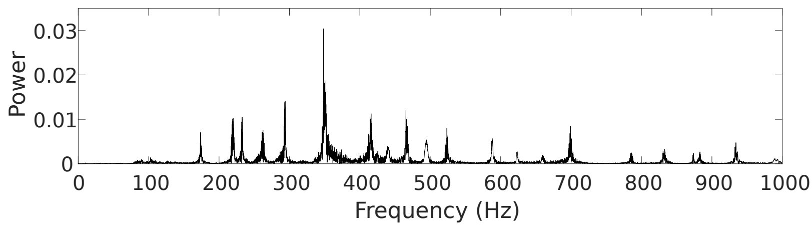

Fig. 3: Frequency-domain plot of the piano signal.

By studying Fig. 3 we might conclude that the most dominating frequency is 350 Hz. Since this frequency corresponds the fundamental frequency of F4 we might further deduce that the key of the music piece is F (major or minor) which is correct. Hence, analyzing $f(x)$ and $\hat{f}(\xi)$ separately we can determine some global properties of the music (such as tempo and key) but we are unable to say anything about its melody. In some sense, this problem is counterintuitive. Even though $f(x)$ and $\hat{f}(\xi)$ contain all possible information of the signal, neither representation provides the relevant information. For this reason we want to construct a TF representation which imitates the musical score and provides a description of the frequencies as a function of time. The connection between TF analysis and music has been studied by several authors [2, 4, 29, 92] and has applications within areas such as transposition [32], transcription [76], beat tracking [89], and compression [57]. Our notation is mainly inspired by the book of Gröchenig [55], which provides an introduction to TF analysis in the framework of Gabor theory.

In order to obtain local TF information we apply a smooth cutoff function g called a window function. Multiplying the signal and the window function we obtain a segment of the signal. If we then take the Fourier transform of this segment we obtain a description of the frequencies of the signal occurring within the segment. This procedure is known as the short-time Fourier transform (STFT) [1, 8, 55]. Formally, the STFT of $f \in L^2(\mathbb{R}^d)$ with respect to

$g \in L^2(\mathbb{R}^d) \setminus \{0\}$ is defined as

$$V_g f(x,\xi) = \int_{\mathbb{R}^d} f(t)\overline{g(t-x)}e^{-2\pi i t\cdot\xi}dt, \quad x,\xi \in \mathbb{R}^d.$$

For practical purposes we want to consider a discretization of the STFT and this is where Gabor frames come into play.

4.1 Stationary Gabor analysis

The connection between the STFT and Gabor frames is straightforward. Let $f \in L^2(\mathbb{R}^d)$ and assume $\{g_{m,n}\}_{m,n\in\mathbb{Z}^d}$ is a stationary Gabor frame for $L^2(\mathbb{R}^d)$. We can rewrite the frame coefficients of f as

$$\langle f, g_{m,n}\rangle = \langle f, M_{mb}T_{na}g\rangle = V_g f(na, mb), \quad m,n \in \mathbb{Z}^d.$$

Hence, the frame coefficients are samples of the STFT along the lattice $a\mathbb{Z}^d \times b\mathbb{Z}^d$ of the TF plane. Likewise we may write the Gabor expansion of f as

$$f = \sum_{m,n\in\mathbb{Z}^d} V_g f(na,mb)\tilde{g}_{m,n}.$$

The TF resolution provided by the STFT depends crucially on the choice of window function. To obtain a good time resolution the window function needs to be well localized in time. As a first attempt, one might therefore choose $g = \chi_{[0,1]^d}$, with $\chi_{[0,1]^d}$ denoting the indicator function on the cube $[0,1]^d$. The Fourier transform of g is then given by

$$\hat{g}(\xi) = \int_{[0,1]^d} e^{-2\pi i x\cdot\xi}dx = \prod_{k=1}^{d} \frac{1-e^{-2\pi i\xi_k}}{2\pi i\xi_k}.$$

With $d = 1$ we get the plots shown in Fig. 4.

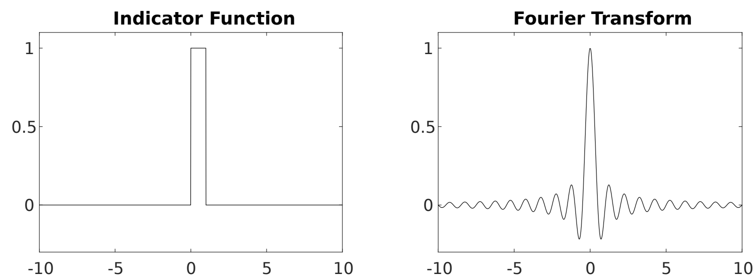

Fig. 4: Plots of $g = \chi_{[0,1]}$ and (the real part of) $\hat{g}$.

The plot of $\hat{g}$ in Fig. 4 reveals that the Fourier transform of $g = \chi_{[0,1]^d}$ decays slowly. This produces a new problem for our analysis as the STFT then provides a poor frequency resolution. One way of realizing this is the following: Using the facts that $\widehat{T_x f} = M_{-x}\hat{f}$ and $\widehat{M_\xi f} = T_\xi \hat{f}$, we can apply Plancherel's theorem to obtain the identity

$$\begin{aligned} V_g f(x,\xi) &= \langle f, M_\xi T_x g\rangle = \langle \hat{f}, T_\xi M_{-x}\hat{g}\rangle \\ &= \langle \hat{f}, e^{2\pi i x\cdot\xi} M_{-x} T_\xi \hat{g}\rangle = e^{-2\pi i x\cdot\xi} V_{\hat{g}}\hat{f}(\xi,-x), \quad f \in L^2(\mathbb{R}^d). \end{aligned} \tag{3}$$

This identity is often referred to as the fundamental identity of TF analysis [55]. We can think of the expression on the right hand-side of (3) as a rotation of the TF plane by 90 degrees. It follows that if $\hat{g}$ decays slowly then the frequency resolution of the STFT becomes poor. Therefore, $g = \chi_{[0,1]^d}$ is not well suited as window function. It turns out that the problem noticed for g is due to a fundamental property of functions known as the uncertainty principle.

The uncertainty principle

The uncertainty principle states roughly that g and $\hat{g}$ cannot both be supported on arbitrary small sets. The exact formulation of the principle can be stated in many different forms [51, 59] and we have chosen to present the formulation of Donoho and Stark [28].

We say that $f \in L^2(\mathbb{R}^d)$ is $\epsilon-$concentrated on a measurable set $T \subseteq \mathbb{R}^d$ if $\|\chi_{T^c} f\|_2 \leq \epsilon \|f\|_2$ with T^c denoting the complement of T. In particular, if $\epsilon \in [0, 1/2)$ then T is called the essential support of f. We note that if $\epsilon = 0$ then $\operatorname{supp}(f) \subseteq T$. Assume $f \in L^2(\mathbb{R}^d) \setminus \{0\}$ is ϵ_T-concentrated on $T \subseteq \mathbb{R}^d$ and $\hat{f}$ is $\epsilon_\Omega-$concentrated on $\Omega \subseteq \mathbb{R}^d$. The uncertainty principle then states

$$|T|\,|\Omega| \geq (1 - \epsilon_T - \epsilon_\Omega)^2.$$

As a corollary, if $\operatorname{supp}(f) \subseteq T$ and $\operatorname{supp}(\hat{f}) \subseteq \Omega$ then $|T||\Omega| \geq 1$. Under these additional assumptions, another version of the uncertainty principle [3, 10] further states that $\infty > |T||\Omega|$ implies $f = 0$.

The uncertainty principle has a particularly interesting consequence for Gabor frames known as the Balian-Low theorem [7, 84]. This result reveals a crucial disadvantage of sampling at the critical density, i.e. $ab = 1$, and thus explains why frames are more appropriate than orthonormal bases for Gabor analysis. Let $W_0(\mathbb{R}^d)$ denote the Wiener space [37, 60, 107] of continuous functions $h \in L^\infty(\mathbb{R}^d)$ with $\sum_{n\in\mathbb{Z}^d} \|h \cdot T_n \chi_{[0,1]^d}\|_\infty < \infty$. The Balian-Low theorem states that if $\{M_{ma^{-1}} T_{na} g\}_{m,n\in\mathbb{Z}^d}$ is a Gabor frame (and thus a Riesz basis) for $L^2(\mathbb{R}^d)$ then both $g \notin W_0(\mathbb{R}^d)$ and $\hat{g} \notin W_0(\mathbb{R}^d)$ [9, 62]. Hence, it is not possible to construct a Gabor frame, sampled at the critical density, using a window function that is well localized in both time and frequency.

The uncertainty principle implies that the idea of simultaneous time and frequency information is unobtainable. It is not possible to construct an ideal TF representation which contains exact information of the frequencies occurring at each time instant. In practice one instead uses a window function such that both g and $\hat{g}$ are decaying rapidly, for instance a Schwartz function $g \in \mathcal{S}(\mathbb{R}^d)$. In particular, one can choose the Gaussian $g(x) = e^{-\|x\|_2^2/2} \in \mathcal{S}(\mathbb{R}^d)$ as shown in Fig. 5.

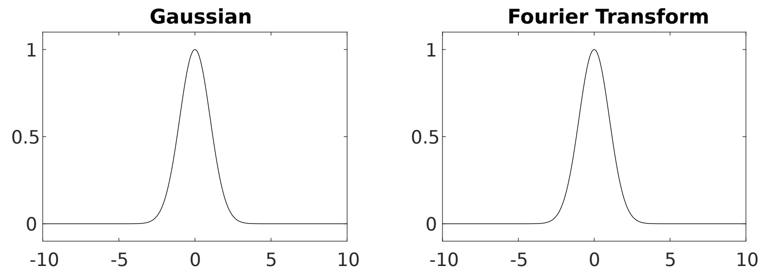

Fig. 5: Plots of $g(x) = e^{-x^2/2}$ (to the left) and $\hat{g}(x) = \sqrt{2\pi}e^{-2\pi^2x^2}$ (to the right).

The Fourier transform of a Gaussian $\varphi_a(x) := e^{-\pi\|x\|_2^2/a}$, with $a > 0$, is simply given by another Gaussian $\widehat{\varphi_a}(\xi) = a^{d/2}\varphi_{1/a}(\xi)$, see [55]. The Gaussian also posses the important property that it minimizes the uncertainty in the Heisenberg-Pauli-Weyl inequality (a classical formulation of the uncertainty principle) [15]. As the resolution of the STFT depends on the choice of window function, the study of suitable window classes are of great importance [36, 55, 79]. However, it is outside the scope of this introduction to go into further details.

The Gabor expansion as a sum of building blocks

In this section we take a more intuitive approach for understanding the (stationary) Gabor expansions

$$f = \sum_{m,n \in \mathbb{Z}^d} V_g f(na, mb)\tilde{g}_{m,n}, \quad f \in L^2(\mathbb{R}^d). \tag{4}$$

We will assume that the window function is a Schwartz function $g \in \mathcal{S}(\mathbb{R}^d)$ such that both g and $\hat{g}$ decay rapidly. The dual window $\tilde{g} = S^{-1}g$ does not necessarily have the same shape as g [44, 73, 104] but it can be shown that

$$\lim_{(a,b)\to(0,0)} \frac{1}{ab}\tilde{g} = g.$$

This result was originally proved by Feichtinger and Zimmermann in [49]. Hence, as the sampling density increases the dual window starts to resemble the original window. This is illustrated in Fig. 6 for the special case when g is a Gaussian.

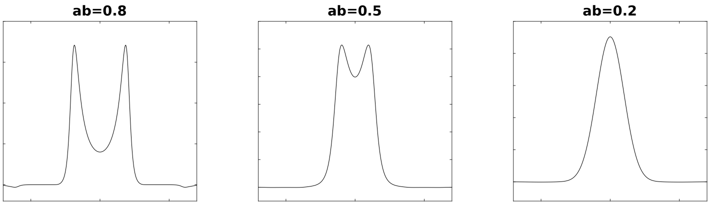

Fig. 6: Dual windows of the Gaussian with increasing density.

Let us now analyze the structure of the dual frame $\{\tilde{g}_{m,n}\}_{m,n}$ by investigating the actions of $M_{mb}T_{na}$ on $\tilde{g}$. In Fig. 7 we have plotted a translated and (the real part of) a modulated version of the Gaussian.

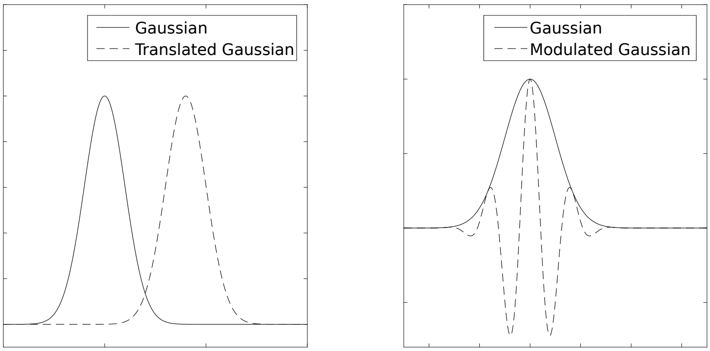

Fig. 7: Translation and (the real part of) modulation of the Gaussian.

Since translation T_{na} corresponds to a horizontal displacement of the function we often refer to this operator as a time shift. On the other hand, since $\widehat{M_\xi f} = T_\xi \hat{f}$ we refer to modulation as a frequency shift. The composition $M_{mb}T_{na}$ is called a TF shift and corresponds to a displacement by na in the horizontal direction and mb in the vertical direction of the TF plane. Hence, we can think of the dual frame $\{M_{mb}T_{na}\tilde{g}\}_{m,n}$ as horizontal and vertical displacements of $\tilde{g}$ in the TF plane [31]. This is illustrated in Fig. 8.

The Gabor expansion in (4) can therefore be interpreted as a weighted sum of building blocks $\{\tilde{g}_{m,n}\}_{m,n}$ centered at the lattice points $a\mathbb{Z}^d \times b\mathbb{Z}^d$ in the TF plane. The weights are samples of the STFT $\{V_g f(na, mb)\}_{m,n}$ determining the contribution of the building blocks to the signal [29]. This point of view can also be found in the more general theory of atomic decompositions developed by Feichtinger and Gröchenig [42, 45, 46].

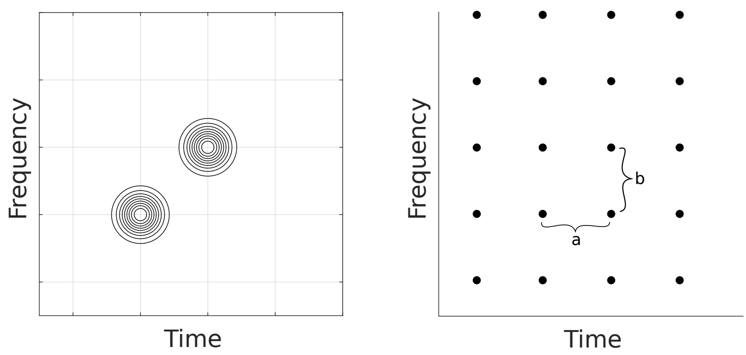

Fig. 8: TF shift $M_b T_a$ of $\tilde{g}$ in the TF plane.

The spectrogram

Based on the TF information contained in the (complex valued) Gabor coefficients $\{V_g f(na, mb)\}_{m,n}$ we can visualize the TF contents of f by plotting the values of $\{|V_g f(na, mb)|^2\}_{m,n}$ in the TF plane. This produces a TF representation called a spectrogram [86]. In Fig. 9 we have plotted a spectrogram for the piano music associated with the musical score in Fig. 1.

To produce such a spectrogram one has to go to the finite settings as the computer can only process vectors of finite lengths. It is outside the scope of this introduction to provide such details and we refer the reader to [91, 101, 103]. In paper D of this thesis we also give a brief summary of Gabor theory in the finite settings. For the associated MATLAB implementation we refer the reader to the toolboxes by Holighaus [6] and Søndergaard [94].

The spectrogram in Fig. 9 has a redundancy of four meaning there are four times as many Gabor coefficients as signal samples. The 22 vertical lines in the spectrogram correspond to the onsets of the music piece and the horizontal lines correspond to the frequencies of the fundamental frequencies and the overtones. The lengths of the horizontal lines correspond to the durations of the notes. We recall that the first note is F4 with a fundamental frequency of 350 Hz and overtones of frequencies 700 Hz, 1050 Hz, 1400 Hz, which is reflected in the spectrogram. By varying the length of the window function we can change the resolution of the spectrogram. A shorter window corresponds to an improved time resolution (and worsened frequency resolution) whereas a longer window corresponds to an improved frequency resolution (and worsened time resolution), see Fig. 10.

The optimal window size depends on the application. For instance, a short window might be preferable for determining the tempo whereas a

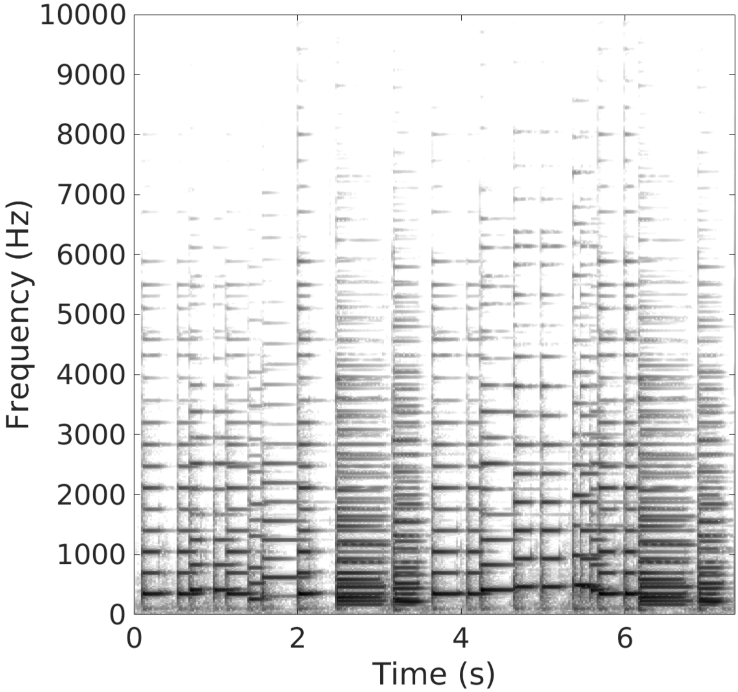

Fig. 9: Spectrogram of the piano music from Fig. 1.

longer window might be better suited for estimating the fundamental frequencies [83]. Once the window function and the lattice parameters are chosen, the spectrogram provides a TF resolution which is independent of the signal under consideration. This stationarity can be seen both as an advantage and a disadvantage of Gabor frames. On the one hand, the stationarity implies a fast and easy to handle implementation [101]. On the other hand, it is not possible to assign different TF resolutions to different regions in the TF plane [31]. In particular, it is not possible for the TF resolution to adapt to the particular characteristics of the signal. The idea behind NSGFs (and the more general multi-window Gabor frames) is to allow the usage of several different window functions to provide a more flexible TF resolution.

4.2 Nonstationary Gabor analysis

As described in Section 3, NSGFs can be implemented in either the time domain or the frequency domain. In other words, the TF resolution obtained through NSGFs can vary along either the time or the frequency axis [6]. This is a limitation compared to general multi-window Gabor frames where a

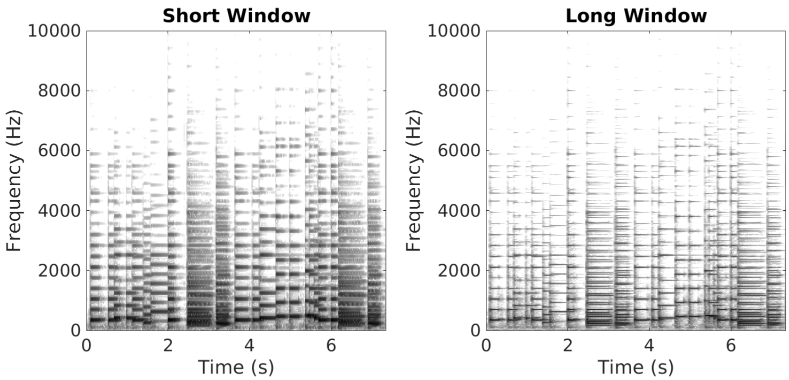

Fig. 10: Spectrograms with different window functions.

particular TF resolution can be assigned to any given region of the TF plane [30]. However, the structure of NSGFs implies a fast implementation based on the FFT which is not the case for general multi-window Gabor frames [83]. In the next two sections we present two practical implementations of NSGFs, one in the time domain and one in the frequency domain. Both implementations were presented in [6] and included in the associated Matlab toolbox.

Scale frames

We first consider an implementation of NSGFs in the time domain. We recall that the atoms are given by

$$g_{m,n}(x) = M_{mb_n} g_n(x) = g_n(x) e^{2\pi i m b_n \cdot x}, \quad m, n \in \mathbb{Z}^d,$$

with $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ and $\{b_n\}_{n\in\mathbb{Z}^d} \subset \mathbb{R}_+$. For a fixed time point n, the associated frequency sampling parameter b_n determines the distance between the vertical sampling points in the TF plane. It is this regularity that allows for an implementation based on the FFT [6]. Hence, the sampling grid associated with a NSGF in the time domain is irregular over time but regular over frequency at each fixed time point as shown in Fig. 11.

The algorithm presented in [6] is based on the idea of applying short window functions around the onsets of a music piece and longer window functions between the onsets. The onsets are estimated using a spectral flux algorithm which applies a preliminary Gabor transform and calculates the sum of (positive) change in magnitude for all frequency bins [26]. The space between two onsets is spanned in such a way that the window length first increases (as we move away from the first onset) and then decreases (as we approach the second onset). More precisely, with $g \in C_c^\infty([0,1])$ denoting a

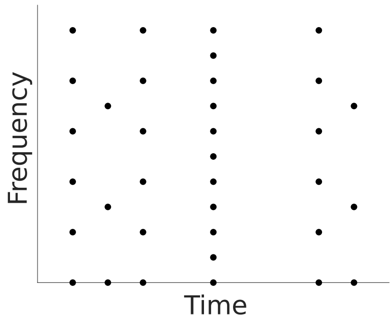

Fig. 11: Sampling grid associated with a NSGF in the time domain.

fixed prototype function, the window functions are given by the dilations

$$g_n(x) := \frac{1}{\sqrt{2^{s_n}}} g\left(\frac{x}{2^{s_n}}\right), \quad n \in \mathbb{Z}.$$

The parameters $\{s_n\}_{n\in\mathbb{Z}}$ constitute an integer-valued scale sequence satisfying $|s_n - s_{n-1}| \in \{0, 1\}$. In this way, adjacent window functions are either of the same length or one is twice as long as the other. To obtain a low redundancy and stable reconstruction, the overlap between adjacent window functions is chosen as $1/3$ of the length for equal windows and $2/3$ of the length of the shorter window for different windows. This guarantees a non-zero overlap between adjacent window function and at most two non-zero windows for any fixed time point. Finally, the numbers of frequency channels are chosen such that the resulting system constitutes a painless NSGF (cf. Section 3.1). Such a NSGF is called a scale frame [6] to emphasize the dependency on the scale sequence $\{s_n\}_{n\in\mathbb{Z}}$. By construction, scale frames apply a redundancy of $\approx 5/3$. In Fig. 12 we have plotted a spectrogram, based on a scale frame, for the piano music in Fig. 1.

We note from Fig. 12 how the adaptive behavior of the scale frame produces a spectrogram which is significantly different from the one in Fig. 9. Between each pair of onsets, the horizontal lines become very thin as a consequence of the improved frequency resolution. On the other hand, the vertical lines associated with the onsets are very sharp due to the good time resolution. Considering the scale frame applies less than half the number of coefficients as used in Fig. 9, the result is very impressive. The reason for this impressive performance is do to the characteristics of the signal. Besides for the chords at the end of bar 2 and bar 4 (cf. Fig. 1), only one note is played at a time and the adaptation procedure therefore correctly detects and separates all 22 onsets. However, for a more complicated music piece (for instance a piano piece with different melodies played by the left and the right hand), the

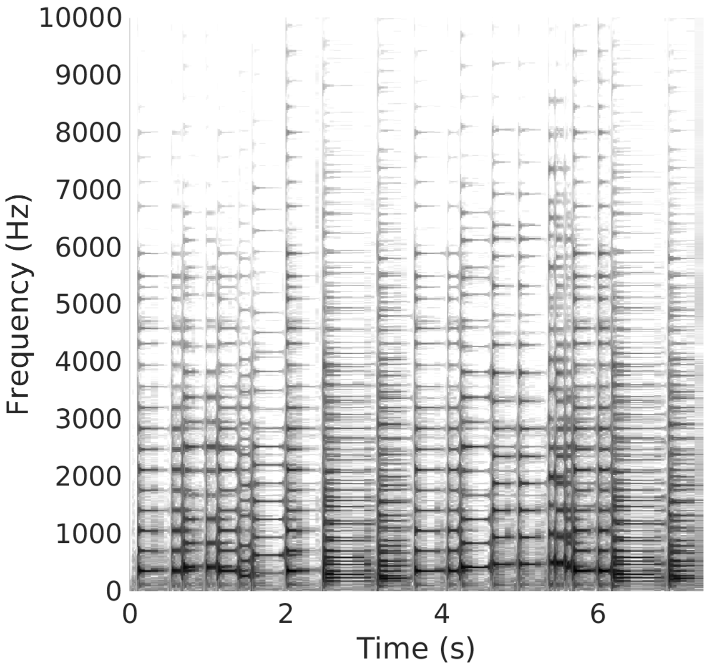

Fig. 12: Spectrogram based on a NSGF in the time domain.

onset detection algorithm will have to choose the most "significant" onsets. For onsets not detected by the algorithm, the associated time resolution will be poor which makes it difficult to determine the correct time instants. We consider this problem in more depth in paper D of this thesis.

The constant Q-transform

In this section we consider an implementation of NSGFs in the frequency domain known as the constant Q-transform [6, 11]. We recall the atoms of the frame are given by

$$h_{m,n}(x) = T_{na_m} h_m(x) = h_m(x - na_m), \quad m, n \in \mathbb{Z}^d,$$

with $\{h_m\}_{m\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ and $\{a_m\}_{m\in\mathbb{Z}^d} \subset \mathbb{R}_+$. The associated sampling grid is irregular over frequency but regular over time for each fixed frequency channel as shown in Fig. 13.

At any given point in the TF plane we can define the $Q-$factor by

$$Q\text{-factor} := \frac{\text{center frequency}}{\text{windows bandwidth}}$$

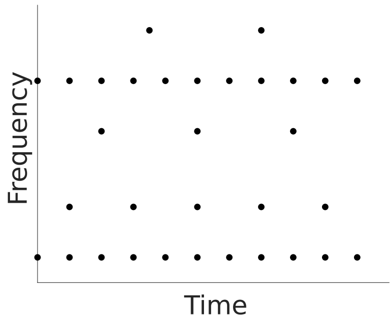

Fig. 13: Sampling grid associated with a NSGF in the frequency domain.

For a stationary Gabor expansion, the $Q-$factor is constant along time but increases with frequency. The idea behind the constant Q-transform is to keep the $Q-$factor constant over the entire TF plane. To do so one needs to apply a NSGF in the frequency domain such that the bandwidths of the window functions increase with frequency. This construction results in a TF representation with good frequency resolution at the lower frequencies and good time resolution at the higher frequencies, similarly to wavelets [17, 58, 87]. Such a resolution might be preferable for music signals since it facilitates the identification of the fundamental frequencies while keeping a good time resolution at higher frequencies for determining the locations of the onsets. The constant $Q-$transform was originally suggested by Brown [11] (in a less efficient form) and implemented in the framework of NSGFs in [6]. The efficient implementation presented in [6] has subsequently been used for various practical applications such as beat tracking and transposition of music signals [32, 66, 67]. In Fig. 14 we have plotted a spectrogram (with dB scaled frequency axis), based on a constant $Q-$transform, for the piano music in Fig. 1.

The spectrogram in Fig. 14 has a redundancy of ≈ 4 and thus applies the same number of coefficients as the spectrogram in Fig. 9. We note the "Christmas tree"-shape of the vertical lines resulting from the varying frequency resolution. The fundamental frequencies are easily accessible and the locations of the associated onsets can be found by looking at the higher frequencies. However, just as for scale frames, the impressive performance is due to the simplistic structure of the music piece. For a more advanced piano piece, the task of determining the locations of the onsets is not at all trivial. Finally, let us note that the constant $Q-$transform is actually a stationary transform as the TF resolution is independent of the signal under consideration. Hence, the terminology of a NSGF is slightly misleading in this case as the TF resolution follows a predetermined rule just as for stationary Gabor

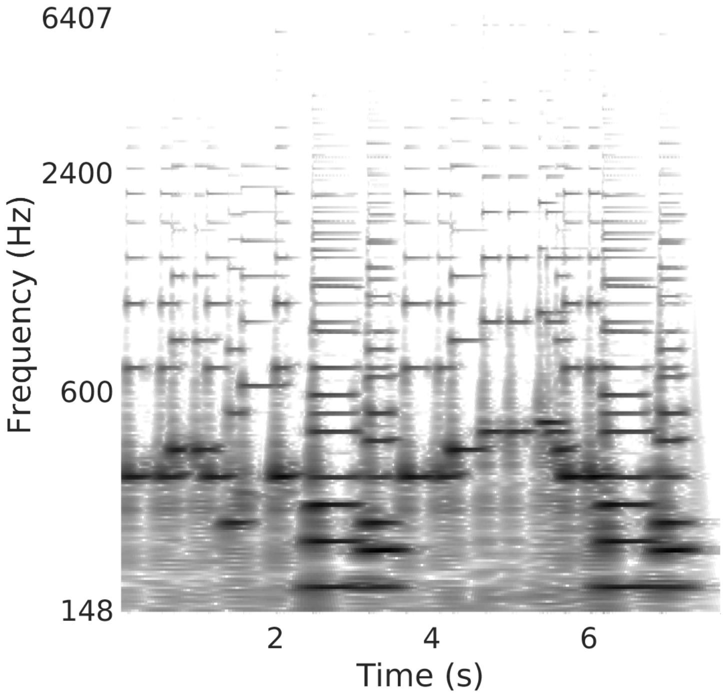

Fig. 14: Spectrogram based on a NSGF in the frequency domain (dB scaled frequency axis).

frames.

We have presented three different approaches for designing TF representations of music signals. The stationary Gabor expansion, scale frames and the constant $Q-$transform. Each TF representation comes with its own advantages and disadvantages and the best choice of TF representation usually depends on the characteristics of the particular signal and the application. As seen from the spectrograms in Fig. 9, Fig. 12, and Fig. 14, music signals tend to produce structured TF representations. This is because music signals are generated from components which are sparse in either time or frequency [93], thereby allowing for horizontal and vertical structures of the spectrograms. We call such spectrograms sparse, since many of the coefficients are close to zero. The sparseness property allows for powerful approximations that with only few non-zero coefficients (compared to the signal length) can produce small reconstruction errors with almost no audible artifacts. In the next section we describe this application in more detail.

5 Nonlinear approximation with frames

Let $\mathcal{D} := \{g_k\}_{k\in\mathbb{N}}$ be a frame for $\mathcal{H}$ with $\|g_k\|_{\mathcal{H}} = 1$ for all $k \in \mathbb{N}$. In approximation theory, such a frame is often referred to as a dictionary for $\mathcal{H}$. Given a possible complicated function $f \in H$ we want to approximate f using linear combinations of the simpler functions from $\mathcal{D}$. We call f the target function and the members of $\mathrm{span}\{g_k\}_{k\in\mathbb{N}}$ the approximants. We define the set of all linear combinations of at most $m \in \mathbb{N}$ elements from $\mathcal{D}$ by

$$\Sigma_m := \left\{ \sum_{k\in\Delta} c_k g_k \;\middle|\; \Delta \subset \mathbb{N},\ \mathrm{card}(\Delta) \le m \right\}.$$

In general, a sum of two elements from Σ_m will need $2m$ terms in its representation by $\{g_k\}_{k\in\mathbb{N}}$. Therefore, the set Σ_m is nonlinear and the associated approximation theory is known as nonlinear approximation [23, 24, 53]. The procedure of approximating f with members of Σ_m is called $m-$term approximation with m measuring the complexity of the approximation [90, 98]. The approximation error associated with Σ_m is measured by

$$\sigma_m(f)_{\mathcal{H}} = \inf_{h\in\Sigma_m} \|f-h\|_{\mathcal{H}}, \quad f \in \mathcal{H}.$$

We note that the sequence $\{\sigma_m(f)_{\mathcal{H}}\}_{m\in\mathbb{N}}$ is non-increasing since $\Sigma_m \subseteq \Sigma_{m+1}$ for all $m \in \mathbb{N}$. In other words, we can increase the resolution of the target function by increasing the complexity of the approximants. This trade-off between resolution and complexity is the main study of approximation theory [22]. For practical purposes, we are especially interested in characterizing functions $f \in \mathcal{H}$ with approximation errors decaying as

$$\sigma_m(f)_{\mathcal{H}} \le C m^{-\alpha}, \quad \forall m \in \mathbb{N}, \tag{5}$$

for some $C > 0$. Therefore, we define the approximation spaces

$$\mathcal{A}^\alpha_\tau(\mathcal{H}) := \left\{ f \in \mathcal{H} \;\middle|\; \sum_{m\in\mathbb{N}} \left(m^\alpha \sigma_m(f)_{\mathcal{H}}\right)^\tau \frac{1}{m} < \infty \right\}, \quad \alpha, \tau \in (0,\infty),$$

together with the associated (quasi-)norms

$$\|f\|_{\mathcal{A}^\alpha_\tau(\mathcal{H})} := \left\| \{m^{\alpha-1/\tau}\sigma_m(f)_{\mathcal{H}}\}_{m\in\mathbb{N}} \right\|_{\ell^\tau} + \|f\|_{\mathcal{H}}, \quad f \in \mathcal{A}^\alpha_\tau(\mathcal{H}).$$

We note that $f \in \mathcal{A}^\alpha_\tau(\mathcal{H})$ is a slightly stronger condition than (5). We also note that the approximation spaces $\mathcal{A}^\alpha_\tau(\mathcal{H})$ are "large" function spaces, in particular they contain all the spaces $\{\Sigma_m\}_{m\in\mathbb{N}}$. One of the goals of nonlinear approximation theory is to characterize the elements of $\mathcal{A}^\alpha_\tau(\mathcal{H})$, thereby providing a description of the elements $f \in \mathcal{H}$ which permit good $m-$term approximations with respect to $\mathcal{D}$ [24].

If $\mathcal{D}$ forms an orthonormal basis for $\mathcal{H}$ then every $f \in \mathcal{H}$ has a unique expansion

$$f = \sum_{k \in \mathbb{N}} \langle f, g_k \rangle g_k, \quad \text{with} \quad \|f\|_{\mathcal{H}}^2 = \sum_{k \in \mathbb{N}} |\langle f, g_k \rangle|^2 .$$

From this one deduces that the best $m-$term approximation to f is obtained by choosing an index set Δ_m corresponding to m terms in the expansion for which $|\langle f, g_k \rangle|$ is largest. The associated approximation error is given by $\sigma_m(f)_{\mathcal{H}} = (\sum_{k \notin \Delta_m} | \langle f, g_k \rangle |^2)^{1/2}$. For this special case have a complete characterization of the approximation space

$$f \in \mathcal{A}_\tau^\alpha(\mathcal{H}) \Leftrightarrow \|\{\langle f, g_k \rangle\}_{k \in \mathbb{N}}\|_{\ell^\tau} < \infty,$$

with $\alpha \in (0, \infty)$ and $0 < \tau := (\alpha + 1/2)^{-1} < 2$. This characterization was proved by Stechkin [102] for the case $\alpha = 1/2$ and for general α by DeVore and Temlyakov [25].

When $\mathcal{D}$ is a redundant frame it becomes much more difficult to provide a complete characterization of $\mathcal{A}_\tau^\alpha(\mathcal{H})$. It is often possible to construct a simpler space $\mathcal{K}$ with $\mathcal{K} \hookrightarrow \mathcal{A}_\tau^\alpha(\mathcal{H})$, i.e., $\mathcal{K} \subset \mathcal{A}_\tau^\alpha(\mathcal{H})$ and $\|f\|_{\mathcal{A}_\tau^\alpha(\mathcal{H})} \leq C \|f\|_{\mathcal{K}}$ for some $C > 0$ and all $f \in \mathcal{K}$, see [53]. Unfortunately, the converse embedding is in general very hard to prove for redundant dictionaries [54]. We call $\mathcal{K} \hookrightarrow \mathcal{A}_\tau^\alpha(\mathcal{H})$ a Jackson embedding and $\mathcal{A}_\tau^\alpha(\mathcal{H}) \hookrightarrow \mathcal{K}$ a Bernstein embedding [12]. The space $\mathcal{K}$ is referred to as a smoothness or sparsity space. We note that a Jackson embedding provides an upper bound on the approximation error $\{\sigma_m(f)_{\mathcal{H}}\}_{m \in \mathbb{N}}$ for all $f \in \mathcal{K}$ whereas a Bernstein embedding provides a lower bound.

An important result by Gröchenig and Samarah [57] shows that modulation spaces can be used as smoothness spaces for stationary Gabor frames. Modulation spaces were introduced by Feichtinger [38] in 1983 and further studied in [40, 43, 44]. Given $1 \leq p < \infty$ and $\gamma \in \mathcal{S}(\mathbb{R}^d) \setminus \{0\}$, the modulation space M^p is defined as those $f \in \mathcal{S}'(\mathbb{R}^d)$ satisfying

$$\|f\|_{M^p} := \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_\gamma f(x, \xi)|^p \, dx d\xi \right)^{1/p} < \infty.$$

It can be shown that M^p is independent of the particular choice of window function γ and different choices yield equivalent norms [38, 55]. For a given Gabor frame $\{g_{m,n}\}_{m,n \in \mathbb{Z}^d} = \{M_{mb} T_{na} g\}_{m,n \in \mathbb{Z}^d}$, the result by Gröchenig and Samarah states that for $1 \leq \tau < p < \infty$ then

$$M^\tau \hookrightarrow \mathcal{A}_\tau^\alpha(M^p), \quad \text{with} \quad \alpha = 1/\tau - 1/p.$$

Hence, the approximation error associated with $f \in M^\tau \subseteq M^p$ (cf. [55] for a proof of this embedding) decays as

$$\sigma_m(f)_{M^p} \leq Cm^{-\alpha} \|f\|_{M^\tau}, \quad \forall m \in \mathbb{N}, \tag{6}$$

for some $C > 0$ and $\alpha = 1/\tau - 1/p$. To illustrate the applications of the theory we return to the piano music from Fig. 1. Since music signals are continuous signals of finite energy, it make sense to consider them in the framework of modulation spaces. The result in (6) therefore indicates that we can expect good approximations of such signals when thresholding the associated Gabor frame expansions. The spectrogram in Fig. 9 is constructed from a Gabor expansion with 655680 Gabor coefficients. Performing hard thresholding and keeping only the 65568 largest coefficients (10% of the total number of coefficients) we obtain the spectrogram in Fig. 15.

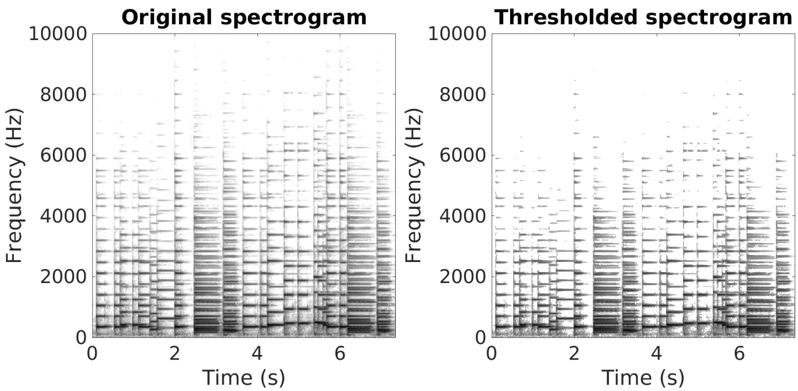

Fig. 15: Original and thresholded spectrogram of the piano music from Fig. 1.

The associated reconstruction error is $\|f - f_{rec}\|_2 / \|f\|_2 \approx 0.005$ and the resulting sound is almost indistinguishable from the original sound. Some of the overtones with high frequencies have been removed, which produces a less colorful timbre of the instrument (one needs good headphones to notice this). One the other hand, some of the low background noises have been suppressed, which results in a more clean sound. In conclusion, with only 10% of the expansion coefficients we obtain an approximation of the signal without almost any decrease in audio quality. It is important to remember that it is the sparsity of the signal which implies such convincing results — for less sparse signals we need more than just 10% of the coefficients for obtaining a reasonable audio quality.

We conclude this introduction by separately addressing each of the four scientific papers and explaining the connection with the theory presented in the introduction.

6 Connection with papers A-D

In Paper A of this thesis we consider nonlinear approximation with general redundant dictionaries (not necessarily restricted to frames). The idea is to generalize the traditional greedy approach [25, 105] for $m-$term approximation (cf. Fig. 15) by including weight functions in the thresholding procedure. For many real world signals the expansions coefficients are correlated and organized in structured sets [77, 78] (see also Fig. 9) and the thresholding procedure should ideally take this dependency into account. The main result of this article is a Jackson embedding for the proposed algorithm under rather general conditions. As an application we generalize the result by Gröchenig and Samarah [57] and provide a numerical comparison between the proposed method and the traditional greedy approach by thresholding music signals expanded in a Gabor dictionary.

In paper B we prove a Jackson embedding for NSGFs in the frequency domain. Whereas modulation spaces works as smoothness spaces for Gabor frames, we need a more general framework for the nonstationary case. Such a framework is provided by decomposition spaces as introduced by Feichtinger and Gröbner [39, 41]. Decomposition spaces are based on a flexible partition of the frequency domain compatible with the structure of NSGFs. The main result of this article is a characterization of the decomposition spaces in terms of the frame coefficients from the NSGFs.

In paper C we consider an approach similar to that of paper B. We use decomposition spaces on the time side to provide a stability result of Hausdorff-Young type for NSGFs in the time domain and prove an associated Jackson inequality for specific choices of parameters. It should be noted that decomposition spaces on the time side are significantly different from decomposition spaces on the frequency side.

Finally, in paper D we consider a practical application and construct a new time-stretching algorithm based on NSGFs in the time domain. Time-stretching is the operation of changing the length of a signal without affecting its frequencies [88]. The paper presents a classical technique known as the phase vocoder [50] in the framework of Gabor theory and extends this techniques to the nonstationary case. The theory is described in the finite settings and the corresponding algorithm is implemented in MATLAB with associated source code and sound files available on-line.

References

[1] J. B. Allen and L. R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558–1564, Nov 1977.

[2] J. F. Alm and J. S. Walker. Time-frequency analysis of musical instruments. *SIAM Rev.*, 44(3):457–476, 2002.

[3] W. O. Amrein and A. M. Berthier. On support properties of L^p-functions and their Fourier transforms. *J. Functional Analysis*, 24(3):258–267, 1977.

[4] G. Assayag, H. G. Feichtinger, and J. F. Rodrigues, editors. *Mathematics and music*. Springer-Verlag, Berlin, 2002.

[5] L. W. Baggett. Processing a radar signal and representations of the discrete Heisenberg group. *Colloq. Math.*, 60/61(1):195–203, 1990.

[6] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco. Theory, implementation and applications of nonstationary Gabor frames. *J. Comput. Appl. Math.*, 236(6):1481–1496, 2011.

[7] R. Balian. Un principe d'incertitude fort en théorie du signal ou en mécanique quantique. *C. R. Acad. Sci. Paris Sér. II Méc. Phys. Chim. Sci. Univers Sci. Terre*, 292(20):1357–1362, 1981.

[8] M. Bastiaans. On the sliding-window representation in digital signal processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):868–873, Aug 1985.

[9] J. J. Benedetto, C. Heil, and D. F. Walnut. Differentiation and the Balian-Low theorem. *J. Fourier Anal. Appl.*, 1(4):355–402, 1995.

[10] M. Benedicks. On Fourier transforms of functions supported on sets of finite Lebesgue measure. *J. Math. Anal. Appl.*, 106(1):180–183, 1985.

[11] J. C. Brown. Calculation of a constant Q spectral transform. *Journal of the Acoustical Society of America*, 89(1):425–434, 1991.

[12] P. L. Butzer and K. Scherer. Jackson and Bernstein-type inequalities for families of commutative operators in Banach spaces. *J. Approximation Theory*, 5:308–342, 1972.

[13] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], second edition, 2016.

[14] C. K. Chui and X. L. Shi. Bessel sequences and affine frames. *Appl. Comput. Harmon. Anal.*, 1(1):29–49, 1993.

[15] M. G. Cowling and J. F. Price. Bandwidth versus time concentration: the Heisenberg-Pauli-Weyl inequality. *SIAM J. Math. Anal.*, 15(1):151–165, 1984.

[16] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory*, 36(5):961–1005, 1990.

[17] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[18] I. Daubechies and A. Grossmann. Frames in the Bargmann space of entire functions. *Comm. Pure Appl. Math.*, 41(2):151–164, 1988.

[19] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5):1271–1283, 1986.

[20] I. Daubechies, H. J. Landau, and Z. Landau. Gabor time-frequency lattices and the Wexler-Raz identity. *J. Fourier Anal. Appl.*, 1(4):437–478, 1995.

[21] N. G. de Bruijn. Uncertainty principles in Fourier analysis. *Inequalities*, pages 57–71, 1967.

[22] R. A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[23] R. A. DeVore. Nonlinear approximation and its applications. In *Multiscale, nonlinear and adaptive approximation*, pages 169–201. Springer, Berlin, 2009.

[24] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[25] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Advances in Computational Mathematics*, 5(1):173–187, Dec 1996.

[26] S. Dixon. Onset detection revisited. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 133–137, Montreal, Quebec, Canada, sep 2006.

[27] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, Nov 2001.

[28] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, 1989.

[29] M. Dörfler. Time-frequency analysis for music signals: A mathematical approach. *Journal of New Music Research*, 30(1):3–12, 2001.

[30] M. Dörfler. Quilted Gabor frames—a new concept for adaptive time-frequency representation. *Adv. in Appl. Math.*, 47(4):668–687, 2011.

[31] M. Dörfler. What time-frequency analysis can do to music signals (and what it can't do...). In *Mathematics and culture. III*, pages 39–51. Springer, Heidelberg, 2012.

[32] M. Dörfler, N. Holighaus, T. Grill, and G. Velasco. Constructing an invertible constant-Q transform with nonstationary Gabor frames. In *n Proceedings of the 14th International Conference on Digital Audio Effects (DAFx 11)*, 2011.

[33] M. Dörfler and E. Matusiak. Nonstationary Gabor frames—existence and construction. *Int. J. Wavelets Multiresolut. Inf. Process.*, 12(3):1450032, 18, 2014.

[34] M. Dörfler and E. Matusiak. Nonstationary Gabor frames—approximately dual frames and reconstruction errors. *Adv. Comput. Math.*, 41(2):293–316, 2015.

[35] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.*, 72:341–366, 1952.

[36] H. G. Feichtinger. On a new Segal algebra. *Monatsh. Math.*, 92(4):269–289, 1981.

[37] H. G. Feichtinger. Banach convolution algebras of Wiener type. In *Functions, series, operators, Vol. I, II (Budapest, 1980)*, volume 35 of *Colloq. Math. Soc. János Bolyai*, pages 509–524. North-Holland, Amsterdam, 1983.

[38] H. G. Feichtinger. Modulation spaces on locally compact abelian groups. Technical report, University of Vienna, 1983.

[39] H. G. Feichtinger. Banach spaces of distributions defined by decomposition methods. II. *Math. Nachr.*, 132:207–237, 1987.

[40] H. G. Feichtinger. Atomic characterizations of modulation spaces through Gabor-type representations. *Rocky Mountain J. Math.*, 19(1):113–125, 1989.

[41] H. G. Feichtinger and P. Gröbner. Banach spaces of distributions defined by decomposition methods. I. *Math. Nachr.*, 123:97–120, 1985.

[42] H. G. Feichtinger and K. Gröchenig. A unified approach to atomic decompositions via integrable group representations. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 52–73. Springer, Berlin, 1988.

[43] H. G. Feichtinger and K. Gröchenig. Gabor wavelets and the Heisenberg group: Gabor expansions and short time Fourier transform from the group theoretical point of view. In *Wavelets*, volume 2 of *Wavelet Anal. Appl.*, pages 359–397. Academic Press, Boston, MA, 1992.

[44] H. G. Feichtinger and K. Gröchenig. Gabor frames and time-frequency analysis of distributions. *J. Funct. Anal.*, 146(2):464–495, 1997.

[45] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. I. *J. Funct. Anal.*, 86(2):307–340, 1989.

[46] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. II. *Monatsh. Math.*, 108(2-3):129–148, 1989.

[47] H. G. Feichtinger and T. Strohmer, editors. *Gabor analysis and algorithms*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 1998. Theory and applications.

[48] H. G. Feichtinger and T. Strohmer, editors. *Advances in Gabor analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2003.

[49] H. G. Feichtinger and G. Zimmermann. A Banach space of test functions for Gabor analysis. In *Gabor analysis and algorithms*, Appl. Numer. Harmon. Anal., pages 123–170. Birkhäuser Boston, Boston, MA, 1998.

[50] J. L. Flanagan and R. M. Golden. Phase vocoder. *The Bell System Technical Journal*, 45(9):1493–1509, Nov 1966.

[51] G. B. Folland and A. Sitaram. The uncertainty principle: a mathematical survey. *J. Fourier Anal. Appl.*, 3(3):207–238, 1997.

[52] D. Gabor. Theory of communication. *J. IEE*, 93(26):429–457, Nov. 1946.

[53] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51–71, 2004.

[54] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. II. Inverse estimates. *Constr. Approx.*, 24(2):157–173, 2006.

[55] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[56] K. Gröchenig. The mystery of Gabor frames. *J. Fourier Anal. Appl.*, 20(4):865–895, 2014.

[57] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–331, 2000.

[58] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15(4):723–736, 1984.

[59] V. Havin and B. Jöricke. *The uncertainty principle in harmonic analysis*, volume 28 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1994.

[60] C. Heil. An introduction to weighted Wiener amalgams. *Wavelets and their Applications (Chennai, January 2002)*, pages 183–216, 2003.

[61] C. Heil. *A basis theory primer*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, expanded edition, 2011.

[62] C. E. Heil. *Wiener amalgam spaces in generalized harmonic analysis and wavelet theory*. PhD thesis, University of Maryland, College Park, 1990.

[63] C. E. Heil and D. F. Walnut. Continuous and discrete wavelet transforms. *SIAM Rev.*, 31(4):628–666, 1989.

[64] E. Hernández, D. Labate, and G. Weiss. A unified characterization of reproducing systems generated by a finite family. II. *J. Geom. Anal.*, 12(4):615–662, 2002.

[65] N. Holighaus. Structure of nonstationary Gabor frames and their dual systems. *Appl. Comput. Harmon. Anal.*, 37(3):442–463, 2014.

[66] N. Holighaus, M. Dörfler, G. A. Velasco, and T. Grill. A framework for invertible, real-time constant-Q transforms. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(4):775–785, April 2013.

[67] A. Holzapfel, G. A. Velasco, N. Holighaus, M. Dörfler, and A. Flexer. Advantages of nonstationary Gabor transforms in beat tracking. In *Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, MIRUM '11, pages 45–50, New York, NY, USA, 2011. ACM.

[68] F. Jaillet. *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore*. PhD thesis, Université d'Aix-Marseille II. Faculté des sciences, 2005.

[69] M. S. Jakobsen and J. Lemvig. Reproducing formulas for generalized translation invariant systems on locally compact abelian groups. *Trans. Amer. Math. Soc.*, 368(12):8447–8480, 2016.

[70] A. J. E. M. Janssen. Gabor representation of generalized functions. *J. Math. Anal. Appl.*, 83(2):377–394, 1981.

[71] A. J. E. M. Janssen. Bargmann transform, Zak transform, and coherent states. *J. Math. Phys.*, 23(5):720–731, 1982.

[72] A. J. E. M. Janssen. Signal analytic proofs of two basic results on lattice expansions. *Appl. Comput. Harmon. Anal.*, 1(4):350–354, 1994.

[73] A. J. E. M. Janssen. Duality and biorthogonality for Weyl-Heisenberg frames. *J. Fourier Anal. Appl.*, 1(4):403–436, 1995.

[74] A. J. E. M. Janssen. The duality condition for Weyl-Heisenberg frames. In *Gabor analysis and algorithms*, Appl. Numer. Harmon. Anal., pages 33–84. Birkhäuser Boston, Boston, MA, 1998.

[75] A. J. E. M. Janssen. Zak transforms with few zeros and the tie. In *Advances in Gabor analysis*, Appl. Numer. Harmon. Anal., pages 31–70. Birkhäuser Boston, Boston, MA, 2003.

[76] A. Klapuri and M. Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, New York, 1st edition, 2006.

[77] M. Kowalski. Sparse regression using mixed norms. *Appl. Comput. Harmon. Anal.*, 27(3):303–324, 2009.

[78] M. Kowalski and B. Torrésani. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. *Signal, Image and Video Processing*, 3(3):251–264, Sep 2009.

[79] W. Kozek. Matched generalized Gabor expansion of nonstationary processes. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 499–503, Nov 1993.

[80] G. Kutyniok and D. Labate. The theory of reproducing systems on locally compact abelian groups. *Colloq. Math.*, 106(2):197–220, 2006.

[81] J. Lemvig. On some Hermite series identities and their applications to Gabor analysis. *Monatsh. Math.*, 182(4):899–912, 2017.

[82] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.

[83] M. Liuni, A. Röbel, E. Matusiak, M. Romito, and X. Rodet. Automatic adaptation of the time-frequency resolution for sound analysis and re-synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5):959–970, May 2013.

[84] F. Low. Complete sets of wave packets. In C. DeTar, editor, *A Passion for Physics - Essay in Honor of Geoffrey Chew*, pages 17–22. World Scientific, Singapore, 1985.

[85] Y. I. Lyubarskiĭ. Frames in the Bargmann space of entire functions. In *Entire and subharmonic functions*, volume 11 of *Adv. Soviet Math.*, pages 167–180. Amer. Math. Soc., Providence, RI, 1992.

[86] S. Mallat. *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition, 2009. The sparse way, With contributions from Gabriel Peyré.

[87] Y. Meyer. *Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1993. Algorithms & applications, Translated from the French and with a foreword by Robert D. Ryan.

[88] E. Moulines and J. Laroche. Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Commun.*, 16(2):175–205, feb 1995.

[89] M. Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, New York, 1st edition, 2015.

[90] K. I. Oskolkov. Polygonal approximation of functions of two variables. *Mat. Sb. (N.S.)*, 107(149)(4):601–612, 639, 1978.

[91] G. E. Pfander. Gabor frames in finite dimensions. In *Finite frames*, Appl. Numer. Harmon. Anal., pages 193–239. Birkhäuser/Springer, New York, 2013.

[92] W. J. Pielemeier, G. H. Wakefield, and M. H. Simoni. Time-frequency analysis of musical signals. *Proceedings of the IEEE*, 84(9):1216–1230, Sep 1996.

[93] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies. Sparse representations in audio and music: From coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005, June 2010.

[94] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs. The Large Time-Frequency Analysis Toolbox 2.0. In M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, editors, *Sound, Music, and Motion*, Lecture Notes in Computer Science, pages 419–442. Springer International Publishing, 2014.

[95] Z. Průša and N. Holighaus. Phase vocoder done right. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 976–980, Aug 2017.

[96] A. Ron and Z. Shen. Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbf{R}^d)$. *Duke Math. J.*, 89(2):237–282, 1997.

[97] A. Ron and Z. Shen. Generalized shift-invariant systems. *Constr. Approx.*, 22(1):1–45, 2005.

[98] E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Math. Ann.*, 63(4):433–476, 1907.

[99] K. Seip. Density theorems for sampling and interpolation in the Bargmann-Fock space. I. *J. Reine Angew. Math.*, 429:91–106, 1992.

[100] K. Seip and R. Wallstén. Density theorems for sampling and interpolation in the Bargmann-Fock space. II. *J. Reine Angew. Math.*, 429:107–113, 1992.

[101] P. L. Søndergaard. *Finite discrete Gabor analysis*. PhD thesis, Technical University of Denmark, 2007.

[102] S. B. Stechkin. On absolute convergence of orthogonal series. *Dokl. Akad. Nauk SSSR (N.S.)*, 102:37–40, 1955.

[103] T. Strohmer. Numerical algorithms for discrete Gabor expansions. In *Gabor analysis and algorithms*, Appl. Numer. Harmon. Anal., pages 267–294. Birkhäuser Boston, Boston, MA, 1998.

[104] T. Strohmer. Approximation of dual Gabor frames, window decay, and wireless communications. *Appl. Comput. Harmon. Anal.*, 11(2):243–262, 2001.

[105] V. N. Temlyakov. Greedy approximation. *Acta Numer.*, 17:235–409, 2008.

[106] D. F. Walnut. Continuity properties of the Gabor frame operator. *J. Math. Anal. Appl.*, 165(2):479–504, 1992.

[107] N. Wiener. Tauberian theorems. *Ann. of Math. (2)*, 33(1):1–100, 1932.

[108] P. J. Wolfe, S. J. Godsill, and M. Dörfler. Multi-Gabor dictionaries for audio time-frequency analysis. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 43–46, 2001.

[109] R. M. Young. *An introduction to nonharmonic Fourier series*, volume 93 of *Pure and Applied Mathematics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.

[110] M. Zibulski and Y. Y. Zeevi. Analysis of multiwindow Gabor-type schemes by frame methods. *Appl. Comput. Harmon. Anal.*, 4(2):188–221, 1997.

Part II

Papers

Paper A

Weighted Thresholding and Nonlinear Approximation

Emil Solsbæk Ottosen and Morten Nielsen

The paper has been submitted to
Indian Journal of Pure And Applied Mathematics

Abstract

We present a new method for performing nonlinear approximation with redundant dictionaries. The method constructs an m−term approximation of the signal by thresholding with respect to a weighted version of its canonical expansion coefficients, thereby accounting for dependency between the coefficients. The main result is an associated strong Jackson embedding, which provides an upper bound on the corresponding reconstruction error. To complement the theoretical results, we compare the proposed method to the pure greedy method and the Windowed-Group Lasso by denoising music signals with elements from a Gabor dictionary.

1 Introduction

Let X be a Banach space equipped with a norm $\|\cdot\|_X$. We consider the problem of approximating a possibly complicated function $f \in X$ using linear combinations of simpler functions $\mathcal{D} := \{g_k\}_{k\in\mathbb{N}}$. We assume $\mathcal{D}$ forms a complete dictionary for X such that $\|g_k\|_X = 1$, for all $k \in \mathbb{N}$, and $\operatorname{span}\{g_k\}_{k\in\mathbb{N}}$ is dense in X. A natural way of performing the approximation is to construct an m-term approximation f_m to f using a linear combination of at most m elements from $\mathcal{D}$ [22, 27, 32]. This leads us to consider the set

$$\Sigma_m(\mathcal{D}) := \left\{ \sum_{k\in\Delta} c_k g_k \;\middle|\; \Delta \subset \mathbb{N},\ \#\Delta \le m \right\}, \quad m \in \mathbb{N}.$$

We note that $\Sigma_m(\mathcal{D})$ is nonlinear since a sum of two elements from $\Sigma_m(\mathcal{D})$ will in general need $2m$ terms in its representation by $\{g_k\}_{k\in\mathbb{N}}$. We measure the approximation error associated to $\Sigma_m(\mathcal{D})$ by

$$\sigma_m(f,\mathcal{D})_X := \inf_{h\in\Sigma_m(\mathcal{D})} \|f-h\|_X, \quad f \in X. \tag{A.1}$$

One of the main challenges of nonlinear approximation theory is to characterize the elements $f \in X$, which have a prescribed rate of approximation $\alpha > 0$ [10, 21, 29]. This is usually done by defining an approximation space $\mathcal{A} \subseteq X$ with the property

$$\sigma_m(f,\mathcal{D})_X = \mathcal{O}\left(m^{-\alpha}\right), \quad \forall f \in \mathcal{A}. \tag{A.2}$$

It is often difficult to characterize the elements of $\mathcal{A}$ directly and a standard approach is therefore to construct a simpler space $\mathcal{K} \subseteq X$ together with a continuous embedding $\mathcal{K} \hookrightarrow \mathcal{A}$. The space $\mathcal{K}$ is referred to as a smoothness or sparseness space and the continuous embedding as a Jackson embedding [2, 11].

In the special case $\{g_k\}_{k\in\mathbb{N}}$ forms an orthonormal basis for a Hilbert space $\mathcal{H}$, it follows from Parseval's identity that the best $m-$term approximation to f is obtained by thresholding the (unique) expansion coefficients and keeping only the m largest coefficients. For a redundant dictionary $\mathcal{D}$, the problem of constructing the best $m-$term approximation is in general computationally intractable [7]. For this reason, various algorithms have been constructed to produce fast and good (but not necessarily the best) $m-$term approximations [6, 12, 19]. By "good approximations" we mean an algorithm $A_m : f \to f_m$ for which there exists an approximation space $\mathcal{T} \subseteq X$ with

$$\|f - f_m\|_X = \mathcal{O}\left(m^{-\alpha}\right), \quad \forall f \in \mathcal{T}.$$

An associated embedding of the type $\mathcal{K} \hookrightarrow \mathcal{T}$ is referred to as a strong Jackson embedding. Whereas a standard Jackson embedding only shows that the error of best m-term approximation decays as in (A.2), a strong Jackson embedding also provides an associated constructive algorithm with this rate of approximation [19].

Traditionally, the expansion coefficients are processed individually with an implicit assumption of independence between the coefficients [12]. However, for many signals such an assumption is not valid as the coefficients are often correlated and organized in structured sets [23, 25]. For instance, it is known that many music signals are generated by components, which are sparse in either time or frequency [30], resulting in sparse and structured time-frequency representations (cf. Fig. A.1 on page 50). In this article we present a thresholding algorithm, which incorporates a weight function to account for dependency between expansion coefficients. The main result is given in Theorem 3.1 and provides a strong Jackson embedding for the proposed method. The idea of exploiting the structure of the expansion coefficients is somewhat similar to the one found in social sparsity [24, 26, 33]. Social sparsity can be seen as a generalization of the classical *Lasso* [38] (also known as basis pursuit denoising [4]) where a weighted neighborhood of each coefficient is considered for deciding whether or not to keep the coefficient. However, whereas social sparsity searches through the dictionary for sparse reconstruction coefficients, the purposed method considers weighted thresholding of the canonical frame coefficients. We compare the proposed algorithm to the Windowed-Group-Lasso (WGL) [25, 34] from social sparsity and the greedy thresholding approach [12, 19] from nonlinear approximation theory by denoising music signals expanded in a Gabor dictionary [5, 20].

The structure of this article is as follows. In Section 2 we introduce the necessary tools from nonlinear approximation theory and in Section 3 we present the proposed algorithm and prove Theorem 3.1. In Section 4 we provide the link to modulation spaces and Gabor frames and in Section 5 we provide the numerical experiments. Finally, in Section 6 we give the conclusions.

2 Elements from nonlinear approximation theory

In this section we define the concepts from nonlinear approximation theory that we will use throughout this article. We refer the reader to [8, 9, 11, 19] for further details. For $\{a_m\}_{m\in\mathbb{N}} \subset \mathbb{C}$, we let $\{a_m^*\}_{m\in\mathbb{N}}$ denote a non-increasing rearrangement of $\{a_m\}_{m\in\mathbb{N}}$ such that $|a_m^*| \geq |a_{m+1}^*|$ for all $m \in \mathbb{N}$.

Definition 2.1 (Lorentz spaces)**.** Given $\tau \in (0,\infty), q \in (0,\infty]$, we define the Lorentz space ℓ_q^τ as the collection of $\{a_m\}_{m\in\mathbb{N}} \subset \mathbb{C}$ satisfying that

$$\|\{a_m\}_{m\in\mathbb{N}}\|_{\ell_q^\tau} := \begin{cases} \left(\sum_{m=1}^\infty \left[m^{1/\tau}\,|a_m^*|\right]^q \frac{1}{m}\right)^{1/q}, & 0 < q < \infty \\ \sup_{m\geq 1} m^{1/\tau}\,|a_m^*|, & q = \infty \end{cases}$$

is finite.

We note that $\|\cdot\|_{\ell_\tau^\tau} = \|\cdot\|_{\ell^\tau}$ for any $\tau \in (0,\infty)$. The Lorentz spaces are rearrangement invariant (quasi-)Banach spaces (Banach spaces if $1 \leq q \leq \tau < \infty$) satisfying the continuous embeddings

$$\ell_{q_1}^{\tau_1} \hookrightarrow \ell_{q_2}^{\tau_2} \quad \text{if} \quad \tau_1 < \tau_2 \quad \text{or if} \quad \tau_1 = \tau_2 \text{ and } q_1 \leq q_2, \tag{A.3}$$

see [3, 8] for details. Let $\mathcal{D} := \{g_k\}_{k\in\mathbb{N}}$ be a complete dictionary for a Banach space X and define $\{\sigma_m(f,\mathcal{D})_X\}_{m\in\mathbb{N}}$ as in (A.1). We will use the following approximation spaces [11].

Definition 2.2 (Approximation spaces)**.** Given $\alpha \in (0,\infty), q \in (0,\infty]$, we define

$$\mathcal{A}_q^\alpha(\mathcal{D},X) := \left\{ f \in X \;\middle|\; \|f\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)} := \|\{\sigma_m(f,\mathcal{D})_X\}_{m\in\mathbb{N}}\|_{\ell_q^{1/\alpha}} + \|f\|_X < \infty \right\}.$$

The quantity $\|\cdot\|_{\mathcal{A}_q^\alpha(\mathcal{D},X)}$ forms a (quasi-)norm on $\mathcal{A}_q^\alpha(\mathcal{D},X)$, and from (A.3) we immediately obtain the continuous embeddings

$$\mathcal{A}_{q_1}^{\alpha_1}(\mathcal{D},X) \hookrightarrow \mathcal{A}_{q_2}^{\alpha_2}(\mathcal{D},X) \quad \text{if} \quad \alpha_1 > \alpha_2 \quad \text{or if} \quad \alpha_1 = \alpha_2 \text{ and } q_1 \leq q_2.$$

We note the $f \in \mathcal{A}_q^\alpha(\mathcal{D},X)$ implies the decay in (A.2) as desired. Following the approach in [12], we define smoothness spaces as follows.

Definition 2.3 (Smoothness spaces)**.** Given $\tau \in (0,\infty), q \in (0,\infty], M > 0$, let

$$\mathcal{K}_q^\tau(\mathcal{D},X,M) := \mathrm{clos}_X \left\{ \sum_{k\in\Delta} c_k g_k \in X \;\middle|\; \Delta \subset \mathbb{N}, \#\Delta < \infty, \|\{c_k\}_{k\in\Delta}\|_{\ell_q^\tau} \leq M \right\}.$$

We then define $\mathcal{K}_q^\tau(\mathcal{D},X) := \cup_{M>0}\mathcal{K}_q^\tau(\mathcal{D},X,M)$ with

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D},X)} := \inf\left\{ M > 0 \;\middle|\; f \in \mathcal{K}_q^\tau(\mathcal{D},X,M) \right\}.$$

It can be shown that $|\cdot|_{\mathcal{K}_q^\tau(\mathcal{D},X)}$ is a (semi-quasi-)norm on $\mathcal{K}_q^\tau(\mathcal{D},X)$ and a (quasi-)norm if $\tau \in (0,1)$.

Remark 2.1. We note that for general $\mathcal{D}$ and X, $f \in \mathcal{K}_q^\tau(\mathcal{D},X)$ does not imply the existence of $\{c_k\}_{k\in\mathbb{N}} \in \ell_q^\tau$ with $f = \sum_{k\in\mathbb{N}} c_k g_k$. All realizations of $\mathcal{K}_q^\tau(\mathcal{D},X)$ considered in this article will, however, guarantee the existence of such reconstruction coefficients (cf. Proposition 2.1 below).

Example 2.1. If $\mathcal{B} = \{g_k\}_{k\in\mathbb{N}}$ is an orthonormal basis for a Hilbert space $\mathcal{H}$, we have the characterization

$$\mathcal{A}_q^\alpha(\mathcal{B},\mathcal{H}) = \mathcal{K}_q^\tau(\mathcal{B},\mathcal{H}) = \left\{ f = \sum_{k\in\mathbb{N}} \langle f, g_k\rangle\, g_k \in \mathcal{H} \,\middle|\, \|\{\langle f, g_k\rangle\}_{k\in\mathbb{N}}\|_{\ell_q^\tau} < \infty \right\},$$

with $\alpha \in (0,\infty), q \in (0,\infty]$ and $0 < \tau = (\alpha + 1/2)^{-1} < 2$ [12, 36]. $\triangle$

Denoting the space of finite sequences on $\mathbb{N}$ by ℓ^0, we define the reconstruction operator $R : \ell^0 \to X$ by

$$R : \{c_k\}_{k\in\Delta} \to \sum_{k\in\Delta} c_k g_k, \quad \{c_k\}_{k\in\Delta} \in \ell^0. \tag{A.4}$$

Recalling that $\|g_k\|_X = 1$, for all $k \in \mathbb{N}$, we can extend this operator to a bounded operator from ℓ^1 to X since

$$\|R\{c_k\}_{k\in\Delta}\|_X \le \sum_{k\in\Delta} |c_k|\, \|g_k\|_X = \|\{c_k\}_{k\in\Delta}\|_{\ell^1}, \quad \forall \{c_k\}_{k\in\Delta} \in \ell^0. \tag{A.5}$$

Following the approach in [19] we introduce the following class of dictionaries.

Definition 2.4 (Hilbertian dictionary)**.** Let $\mathcal{D} = \{g_k\}_{k\in\mathbb{N}}$ be a dictionary in a Banach space X. Given $\tau \in (0,\infty), q \in (0,\infty]$, we say that $\mathcal{D}$ is $\ell_q^\tau-$hilbertian if the reconstruction operator R given in (A.4) is bounded from ℓ_q^τ to X.

It follows from (A.5) and (A.3) that every $\mathcal{D}$ is $\ell_q^\tau-$hilbertian if $\tau < 1$. According to [19, Proposition 3] we have the following characterization.

Proposition 2.1. *Assume $\mathcal{D}$ is ℓ_1^p-hilbertian with $p \in (1,\infty)$. Let $\tau \in (0,p)$ and $q \in (0,\infty]$. For all $f \in \mathcal{K}_q^\tau(\mathcal{D},X)$, there exists some $\boldsymbol{c} := \boldsymbol{c}_{\tau,q}(f) \in \ell_q^\tau$ with $f = R\boldsymbol{c}$ and $\|\boldsymbol{c}\|_{\ell_q^\tau} = |f|_{\mathcal{K}_q^\tau(\mathcal{D},X)}$. If $1 < q \le \tau < \infty$, then $\boldsymbol{c}$ is unique. Consequently*

$$|f|_{\mathcal{K}_q^\tau(\mathcal{D},X)} = \min_{\boldsymbol{c}\in\ell_q^\tau, f=R\boldsymbol{c}} \|\boldsymbol{c}\|_{\ell_q^\tau},$$

and

$$\mathcal{K}_q^\tau(\mathcal{D},X) = \left\{ \sum_{k\in\mathbb{N}} c_k g_k \in X \,\middle|\, \|\{c_k\}_{k\in\mathbb{N}}\|_{\ell_q^\tau} < \infty \right\}$$

is a (quasi-)Banach space with $\mathcal{K}_q^\tau(\mathcal{D},X) \hookrightarrow X$.

In the next section we describe the proposed algorithm using the framework presented in this section.

3 Weighted thresholding

Let $\mathcal{D} := \{g_k\}_{k\in\mathbb{N}}$ be a complete dictionary for a Banach space X and let R be the reconstruction operator defined in (A.4). Given $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$, we let $\pi : \mathbb{N} \to \mathbb{N}$ denote a bijective mapping such that $\{|c_{\pi(k)}|\}_{k\in\mathbb{N}}$ is non-increasing, i.e, $\{c_{\pi(k)}\}_{k\in\mathbb{N}} = \{c_k^*\}_{k\in\mathbb{N}}$. For $f \in X$, $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$, and $m \in \mathbb{N}$, a standard way of constructing an m-term approximant to f from $\mathcal{D}$ is by thresholding

$$f_m := f_m(\pi, \{c_k\}_{k\in\mathbb{N}}, \mathcal{D}) := R\{c_{\pi(k)}\}_{k=1}^{m} = \sum_{k=1}^{m} c_{\pi(k)} g_{\pi(k)}. \tag{A.6}$$

For all practical purpose we choose $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$ as reconstruction coefficients for f such that $f = R\{c_k\}_{k\in\mathbb{N}}$ (cf. Remark 2.1). With this choice, the approximants $\{f_m\}_{m\in\mathbb{N}}$ converge to f as $m \to \infty$. The thresholding approach in (A.6) chooses the m elements from $\mathcal{D}$ corresponding to the m largest of the coefficients $\{c_k\}_{k\in\mathbb{N}}$. As mentioned in the introduction, many real world signals have an inherent structure between the expansion coefficients, which should be accounted for in the approximation procedure. Therefore, we would like to construct an algorithm which preserves local coherence, such that a small coefficient c_1 might be preserved, in exchange for a larger (isolated) coefficient c_2, if c_1 belongs to a neighborhood with many large coefficients. This leads us to consider banded Toeplitz matrices.

Let Λ denote a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$, i.e., the non-zero entries $\Lambda_{(i,j)}$ satisfy $|i-j| \leq \Omega$. We assume the scalar on the diagonal of Λ is positive. Given $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$ we define $\{c_k^\Lambda\}_{k\in\mathbb{N}} := \Lambda(\{|c_k|\}_{k\in\mathbb{N}})$.

Example 3.1. With $\Omega = 2$ we obtain

$$\{c_k^\Lambda\}_{k\in\mathbb{N}} = \begin{bmatrix} \lambda_0 & \lambda_1 & \lambda_2 & 0 & 0 & 0 & \cdots \\ \lambda_{-1} & \lambda_0 & \lambda_1 & \lambda_2 & 0 & 0 & \cdots \\ \lambda_{-2} & \lambda_{-1} & \lambda_0 & \lambda_1 & \lambda_2 & 0 & \cdots \\ 0 & \lambda_{-2} & \lambda_{-1} & \lambda_0 & \lambda_1 & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} |c_1| \\ |c_2| \\ |c_3| \\ |c_4| \\ \vdots \end{bmatrix},$$

with $\lambda_0 > 0$ and $\lambda_l \geq 0$ for all $l \in \{-2, -1, 1, 2\}$. We note that

$$c_k^\Lambda = \lambda_{-2}\,|c_{k-2}| + \cdots + \lambda_0\,|c_k| + \cdots + \lambda_2\,|c_{k+2}| = \sum_{j=k-\Omega}^{k+\Omega} \lambda_{j-k}\,|c_j|,$$

for all $k \in \mathbb{N}$. $\triangle$

Given $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$, we let $\pi_\Lambda : \mathbb{N} \to \mathbb{N}$ denote a bijective mapping such that the sequence $\{c^\Lambda_{\pi_\Lambda(k)}\}_{k\in\mathbb{N}}$ is non-increasing. If Λ is the identity map, we write π instead of π_Λ to be consistent with the notation of (A.6). We shall use the following technical result in the proof of Theorem 3.1.

Lemma 3.1. *Let Λ be a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$ and let $p \in (0,\infty), q \in (0,\infty]$. There exists a constant $C > 0$, such that for any non-increasing sequence $\{c_k\}_{k\in\mathbb{N}} \in \ell^p_q$ we have the estimate*

$$\left\|\{c_{\pi_\Lambda(k)}\}_{k=m+1}^{\infty}\right\|_{\ell^p_q} \leq C\left\|\{c_k\}_{k=m+1-\Omega}^{\infty}\right\|_{\ell^p_q}, \quad \forall m \geq \Omega.$$

Proof. Fix $m \in \mathbb{N}$ and let $\boldsymbol{a} := \{|c_{\pi_\Lambda(k)}|\}_{k=m+1}^{\infty}$ and $\boldsymbol{b} := \{|c_k|\}_{k=m+1}^{\infty}$. If $\boldsymbol{a}$ contains the same coefficients as $\boldsymbol{b}$, then there is nothing to prove. If this is not the case, then there exists $k' \leq m$, with $|c_{k'}| \in \boldsymbol{a}$, and $k'' \geq m+1$, with $|c_{k''}| \notin \boldsymbol{a}$, satisfying

$$\sum_{j=k'-\Omega}^{k'+\Omega} \lambda_{j-k'}\left|c_j\right| \leq \sum_{j=k''-\Omega}^{k''+\Omega} \lambda_{j-k''}\left|c_j\right| \Rightarrow |c_{k'}| \leq \frac{1}{\lambda_0}\sum_{j=k''-\Omega}^{k''+\Omega} \lambda_{j-k''}\left|c_j\right|.$$

Denoting the maximum value of the λ_l's by $\lambda_{\max}$ and the maximum value of the $\left|c_j\right|$'s, for $j \in [k''-\Omega, k''+\Omega]$, by $\left|c_{\tilde{k}}\right|$, we thus get

$$|c_{k'}| \leq \frac{\lambda_{\max}}{\lambda_0}(2\Omega+1)\left|c_{\tilde{k}}\right|.$$

Since $\tilde{k} \in [k''-\Omega, k''+\Omega]$, and $k'' \geq m+1$, then $\tilde{k} \in [m+1-\Omega, \infty)$. We conclude that $c_{\tilde{k}} \in \{c_k\}_{k=m+1-\Omega}^{\infty}$. The lemma follows directly from this observation. □

Given $f \in X$, a set of reconstruction coefficients $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$, and $m \in \mathbb{N}$, we generalize the notation of (A.6) and construct an m-term approximant to f by

$$f_m^\Lambda := f_m^\Lambda(\pi_\Lambda, \{c_k\}_{k\in\mathbb{N}}, \mathcal{D}) := R\{c_{\pi_\Lambda(k)}\}_{k=1}^{m} = \sum_{k=1}^{m} c_{\pi_\Lambda(k)} g_{\pi_\Lambda(k)}. \tag{A.7}$$

If Λ is the identity map, we just obtain the approximant f_m given in (A.6). If not, we obtain an approximant which chooses the elements of $\{g_k\}_{k\in\mathbb{N}}$ corresponding to the indices of the m largest of the weighted coefficients $\{c_k^\Lambda\}_{k\in\mathbb{N}}$. We generalize the notation of [19] and define weighted thresholding approximation spaces as follows.

Definition 3.1. Let Λ be a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$. Given $\alpha \in (0,\infty), q \in (0,\infty]$, we define

$$\mathcal{T}_q^{\alpha}(\mathcal{D}, X, \Lambda) := \left\{ f \in X \;\middle|\; \|f\|_{\mathcal{T}_q^{\alpha}(\mathcal{D},X,\Lambda)} := |f|_{\mathcal{T}_q^{\alpha}(\mathcal{D},X,\Lambda)} + \|f\|_X < \infty \right\},$$

with

$$|f|_{\mathcal{T}_q^{\alpha}(\mathcal{D},X,\Lambda)} := \begin{cases} \inf_{\pi_\Lambda, \{c_k\}_{k\in\mathbb{N}}} \left(\sum_{m=1}^{\infty} \left[m^{\alpha} \left\| f - f_m^{\Lambda} \right\|_X \right]^q \frac{1}{m} \right)^{1/q}, & 0 < q < \infty \\ \inf_{\pi_\Lambda, \{c_k\}_{k\in\mathbb{N}}} \left(\sup_{m\geq 1} m^{\alpha} \left\| f - f_m^{\Lambda} \right\|_X \right), & q = \infty \end{cases}$$

When Λ is the identity mapping, we write $\mathcal{T}_q^{\alpha}(\mathcal{D}, X)$ instead of $\mathcal{T}_q^{\alpha}(\mathcal{D}, X, \Lambda)$.

Remark 3.1. We note that the expression for $|f|_{\mathcal{T}_q^{\alpha}(\mathcal{D},X,\Lambda)}$ cannot be written using the Lorentz norm, since the sequence $\{\|f - f_m^{\Lambda}\|_X\}_{m\in\mathbb{N}}$ might not be non-increasing.

In order to prove Theorem 3.1 we need to impose further assumptions on the dictionary $\mathcal{D}$. Given $\tau \in (0,\infty), q \in (0,\infty]$, we call $\mathcal{D}$ an atomic decomposition (AD) [16, 17] for X, with respect to ℓ_q^{τ}, if there exists a sequence $\{\tilde{g}_k\}_{k\in\mathbb{N}}$, in the dual space X', such that

1. There exist $0 < C' \leq C'' < \infty$ with

$$C' \left\| \{\langle f, \tilde{g}_k \rangle\}_{k\in\mathbb{N}} \right\|_{\ell_q^{\tau}} \leq \|f\|_X \leq C'' \left\| \{\langle f, \tilde{g}_k \rangle\}_{k\in\mathbb{N}} \right\|_{\ell_q^{\tau}}, \quad \forall f \in X.$$

2. The reconstruction operator R given in (A.4) is bounded from ℓ_q^{τ} onto X and we have the expansions

$$R(\{\langle f, \tilde{g}_k \rangle\}_{k\in\mathbb{N}}) = \sum_{k\in\mathbb{N}} \langle f, \tilde{g}_k \rangle\, g_k = f, \quad \forall f \in X.$$

Example 3.2. Standard examples of ADs are Gabor frames for modulation spaces [20] and wavelets for Besov spaces [10]. However, many other examples have been constructed in the general framework of decomposition spaces [1, 14, 15]. For instance curvelets, shearlets and nonstationary Gabor frames (or generalized shift-invariant systems) [1, 28, 40]. $\triangle$

We can now present the main result of this article.

Theorem 3.1. *Let Λ be a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$ and let $p \in (1,\infty)$, $\tau \in (0,p)$, $q \in (0,\infty]$. If $\mathcal{D} = \{g_k\}_{k\in\mathbb{N}}$ is ℓ_1^p−hilbertian and forms an AD for X, with respect to ℓ_q^{τ}, then*

$$\mathcal{K}_q^{\tau}(\mathcal{D}, X) \hookrightarrow \mathcal{T}_q^{\alpha}(\mathcal{D}, X, \Lambda) \hookrightarrow \mathcal{A}_q^{\alpha}(\mathcal{D}, X),$$

with $\alpha = 1/\tau - 1/p > 0$.

Proof. The embedding $\mathcal{T}_q^{\alpha}(\mathcal{D}, X, \Lambda) \hookrightarrow \mathcal{A}_q^{\alpha}(\mathcal{D}, X)$ follows directly from the definitions of these spaces. Let us now show that $\mathcal{K}_q^{\tau}(\mathcal{D}, X) \hookrightarrow \mathcal{T}_q^{\alpha}(\mathcal{D}, X, \Lambda)$. Since the Lorentz spaces are rearrangement invariant, we may assume the canonical coefficients $\boldsymbol{d} := \{d_k\}_{k\in\mathbb{N}} := \{\langle f, \tilde{g}_k\rangle\}_{k\in\mathbb{N}}$ form a non-increasing sequence. Given $f \in \mathcal{K}_q^{\tau}(\mathcal{D}, X)$, the ℓ_1^p−hilbertian property of $\mathcal{D}$ implies

$$\|f\|_X = \|R\boldsymbol{d}\|_X \leq C_1 \|\boldsymbol{d}\|_{\ell_1^p}.$$

Defining $f_m^{\Lambda}(\pi_\Lambda, \boldsymbol{d}, \mathcal{D}) = R\{d_{\pi_\Lambda(k)}\}_{k=1}^m$ as in (A.7), we thus get

$$\begin{aligned}
\|f\|_{\mathcal{T}_q^{\alpha}(\mathcal{D},X,\Lambda)} &= \inf_{\pi_\Lambda, \{c_k\}_{k\in\mathbb{N}}} \left(\sum_{m=1}^{\infty} \left[m^{\alpha} \left\| f - f_m^{\Lambda} \right\|_X \right]^q \frac{1}{m} \right)^{1/q} + \|f\|_X \\
&\leq \left(\sum_{m=1}^{\infty} \left[m^{\alpha} \left\| f - f_m^{\Lambda}(\pi_\Lambda, \boldsymbol{d}, \mathcal{D}) \right\|_X \right]^q \frac{1}{m} \right)^{1/q} + C_1 \|\boldsymbol{d}\|_{\ell_1^p}. \quad \text{(A.8)}
\end{aligned}$$

Now, since

$$\begin{aligned}
\left\| f - f_m^{\Lambda}(\pi_\Lambda, \boldsymbol{d}, \mathcal{D}) \right\|_X &= \left\| R\{d_{\pi_\Lambda(k)}\}_{k=m+1}^{\infty} \right\|_X \\
&\leq C_1 \left\| \{d_{\pi_\Lambda(k)}\}_{k=m+1}^{\infty} \right\|_{\ell_1^p} \leq C_1 \|\boldsymbol{d}\|_{\ell_1^p},
\end{aligned}$$

we get the following estimate for the first Ω terms in (A.8)

$$\sum_{m=1}^{\Omega} \left[m^{\alpha} \left\| f - f_m^{\Lambda}(\pi_\Lambda, \boldsymbol{d}, \mathcal{D}) \right\|_X \right]^q \frac{1}{m} \leq C_2 \|\boldsymbol{d}\|_{\ell_1^p}^q. \quad \text{(A.9)}$$

For $m \geq \Omega + 1$, Lemma 3.1 implies

$$\begin{aligned}
\left\| f - f_m^{\Lambda}(\pi_\Lambda, \boldsymbol{d}, \mathcal{D}) \right\|_X &\leq C_1 \left\| \{d_{\pi_\Lambda(k)}\}_{k=m+1}^{\infty} \right\|_{\ell_1^p} \leq C_3 \left\| \{d_k\}_{k=m+1-\Omega}^{\infty} \right\|_{\ell_1^p} \\
&= C_3 \left\| \boldsymbol{d} - \{d_k\}_{k=1}^{m-\Omega} \right\|_{\ell_1^p} = C_3 \sigma_{m-\Omega}(\boldsymbol{d}, \mathcal{B})_{\ell_1^p},
\end{aligned}$$

with $\mathcal{B}$ denoting the canonical basis of ℓ_1^p. Hence,

$$\begin{aligned}
\sum_{m=\Omega+1}^{\infty} \left[m^{\alpha} \left\| f - f_m^{\Lambda}(\pi_\Lambda, \boldsymbol{d}, \mathcal{D}) \right\|_X \right]^q \frac{1}{m} &\leq C_3 \sum_{m=\Omega+1}^{\infty} \left[m^{\alpha} \sigma_{m-\Omega}(\boldsymbol{d}, \mathcal{B})_{\ell_1^p} \right]^q \frac{1}{m} \\
&= C_3 \sum_{m=1}^{\infty} \frac{\left[(m+\Omega)^{\alpha} \sigma_m(\boldsymbol{d}, \mathcal{B})_{\ell_1^p} \right]^q}{m+\Omega} \\
&\leq C_4 \left\| \{\sigma_m(\boldsymbol{d}, \mathcal{B})_{\ell_1^p}\}_{m\in\mathbb{N}} \right\|_{\ell_q^{1/\alpha}}^q. \quad \text{(A.10)}
\end{aligned}$$

Combining (A.9) and (A.10), then (A.8) yields

$$\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D},X,\Lambda)} \leq C_5 \left(\left\| \{\sigma_m(\boldsymbol{d},\mathcal{B})_{\ell_1^p}\}_{m\in\mathbb{N}} \right\|_{\ell_q^{1/\alpha}} + \|\boldsymbol{d}\|_{\ell_1^p} \right) = C_5 \|\boldsymbol{d}\|_{\mathcal{A}_q^\alpha(\mathcal{B},\ell_1^p)} .$$

Applying [18, Theorem 3.1] thus yields

$$\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D},X,\Lambda)} \leq C_6 \, |\boldsymbol{d}|_{\mathcal{K}_q^\tau(\mathcal{B},\ell_1^p)} = C_6 \|\boldsymbol{d}\|_{\ell_q^\tau} \leq C_7 \|f\|_X . \tag{A.11}$$

Finally, since $\mathcal{D}$ is ℓ_1^p-hilbertian, Proposition 2.1 states that we can find $\boldsymbol{c} \in \ell_q^\tau$ with $f = R\boldsymbol{c}$ and $\|\boldsymbol{c}\|_{\ell_q^\tau} = |f|_{\mathcal{K}_q^\tau(\mathcal{D},X)}$. Combining this with (A.11) we arrive at

$$\|f\|_{\mathcal{T}_q^\alpha(\mathcal{D},X,\Lambda)} \leq C_7 \|R\boldsymbol{c}\|_X \leq C_8 \|\boldsymbol{c}\|_{\ell_q^\tau} = C_8 \, |f|_{\mathcal{K}_q^\tau(\mathcal{D},X)} .$$

This completes the proof. □

Remark 3.2. The assumption in Theorem 3.1 of $\mathcal{D}$ being ℓ_1^p-hilbertian directly implies the boundedness of $R : \ell_q^\tau \to X$ in the definition of an AD since $\tau \in (0,p)$. It should also be noted that if Λ is the identity operator, then Theorem 3.1 holds without the assumption of an AD — this was proven in [19, Theorem 6]. However, in contrast to the proof presented here, there is no constructive way of obtaining the sparse expansion coefficients in the proof of [19, Theorem 6].

The Jackson embedding in Theorem 3.1 is strong in the sense that there is an associated algorithm, which obtains the approximation rate. Given $f \in X$, the algorithm goes as follows:

1. Calculate the canonical coefficients $\{d_k\}_{k\in\mathbb{N}} = \{\langle f, \tilde{g}_k\rangle\}_{k\in\mathbb{N}}$.
2. Construct the weighted coefficients $\{d_k^\Lambda\}_{k\in\mathbb{N}}$ according to Λ by
$$d_k^\Lambda = \sum_{j=k-\Omega}^{k+\Omega} \lambda_{j-k} \left| d_j \right|, \quad k \in \mathbb{N}.$$
3. Choose $\pi_\Lambda : \mathbb{N} \to \mathbb{N}$ such that $\{d^\Lambda_{\pi_\Lambda(k)}\}_{k\in\mathbb{N}}$ is non-increasing.
4. Construct an m-term approximation to f by
$$f_m^\Lambda(\pi_\Lambda, \{d_k\}_{k\in\mathbb{N}}, \mathcal{D}) = \sum_{k=1}^{m} d_{\pi_\Lambda(k)} g_{\pi_\Lambda(k)}.$$

With this construction, Theorem 3.1 states that

$$\left\| f - f_m^\Lambda(\pi_\Lambda, \{d_k\}_{k\in\mathbb{N}}, \mathcal{D}) \right\|_X = \mathcal{O}(m^{-\alpha}), \quad \forall f \in \mathcal{K}_q^\tau(\mathcal{D}, X),$$

with $\alpha = 1/\tau - 1/p > 0$. In the next section we consider the particular case where the dictionary is a Gabor frame and the smoothness space is a modulation space.

4 Modulation spaces and Gabor frames

In this section we choose X as the modulation space M^p [13], with $1 \le p < \infty$, consisting of all $f \in \mathcal{S}'(\mathbb{R}^d)$ satisfying

$$\|f\|_{M^p} := \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |V_\gamma f(x,y)|^p \, dxdy \right)^{1/p} < \infty.$$

Here, $V_\gamma f(x,y)$ denotes the short-time Fourier transform of f with respect to a window function $\gamma \in \mathcal{S}(\mathbb{R}^d) \setminus \{0\}$ (it can be shown that M^p is independent of the particular choice of window function [20]). Given $g \in \mathcal{S}(\mathbb{R}^d) \setminus \{0\}$ and lattice parameters $a, b \in (0, \infty)$, we consider the Gabor system $\{g_{m,n}\}_{m,n \in \mathbb{Z}^d}$ defined by

$$g_{m,n}(t) := g(t - na)e^{2\pi i m b \cdot t}, \quad t \in \mathbb{R}^d.$$

If $\{g_{m,n}\}_{m,n}$ is a frame for $L^2(\mathbb{R}^d)$ (see [5] for details) then there exists a dual frame $\{\tilde{g}_{m,n}\}_{m,n}$ such that $\{g_{m,n}\}_{m,n}$ forms an AD for M^p with respect to ℓ^p for all $1 \le p < \infty$ [20]. We have the following version of [21, Proposition 3].

Proposition 4.1. *Let Λ be a banded non-negative Toeplitz matrix with bandwidth $\Omega \in \mathbb{N}$. Let $1 \le \tau < p < \infty$ and $g \in \mathcal{S}(\mathbb{R}^d) \setminus \{0\}$. If $\mathcal{D} = \{g_{m,n}\}_{m,n \in \mathbb{Z}^d}$ is a Gabor frame for $L^2(\mathbb{R}^d)$, with $\|g_{m,n}\|_{M^p} = 1$ for all $m, n \in \mathbb{Z}^d$, then*

$$M^\tau(\mathbb{R}^d) = \mathcal{K}_\tau^\tau(\mathcal{D}, M^p) \hookrightarrow \mathcal{T}_\tau^\alpha(\Lambda, \mathcal{D}, M^p) \hookrightarrow \mathcal{A}_\tau^\alpha(\mathcal{D}, M^p), \quad \alpha = 1/\tau - 1/p,$$

where the first equality is with equivalent norms.

Proof. We first note that $\mathcal{D}$ simultaneously forms an AD for both $M^p(\mathbb{R}^d)$ and $M^\tau(\mathbb{R}^d)$. Since $\mathcal{D}$ constitutes an AD for $M^p(\mathbb{R}^d)$, and $p > 1$, we get

$$\|R\boldsymbol{c}\|_{M^p} \le C_1 \|\boldsymbol{c}\|_{\ell^p} \le C_2 \|\boldsymbol{c}\|_{\ell_1^p}, \quad \forall \boldsymbol{c} \in \ell_1^p,$$

which shows that $\mathcal{D}$ is a ℓ_1^p-hilbertian dictionary for M^p. Hence, the embeddings in Proposition 4.1 follows from Theorem 3.1. Let us now prove that $M^\tau(\mathbb{R}^d) = \mathcal{K}_\tau^\tau(\mathcal{D}, M^p)$ with equivalent norms. According to Proposition 2.1 then

$$\mathcal{K}_\tau^\tau(\mathcal{D}, M^p) = \left\{ \sum_{m \in \mathbb{Z}^d} \sum_{n \in \mathbb{Z}^d} c_{m,n} g_{m,n} \in M^p \;\middle|\; \left\| \{c_{m,n}\}_{m,n \in \mathbb{Z}^d} \right\|_{\ell^\tau} < \infty \right\} \quad \text{with}$$

$$|f|_{\mathcal{K}_\tau^\tau(\mathcal{D}, M^p)} = \min_{\boldsymbol{c} \in \ell^\tau, f = R\boldsymbol{c}} \|\boldsymbol{c}\|_{\ell^\tau}, \quad f \in \mathcal{K}_\tau^\tau(\mathcal{D}, M^p).$$

Since $\mathcal{D}$ constitutes an AD for $M^\tau(\mathbb{R}^d)$, then for $f \in M^\tau(\mathbb{R}^d)$ we have

$$f = \sum_{m \in \mathbb{Z}^d} \sum_{n \in \mathbb{Z}^d} \langle f, \tilde{g}_{m,n} \rangle \, g_{m,n}, \quad \text{with} \quad \left\| \{\langle f, \tilde{g}_{m,n} \rangle\}_{m,n \in \mathbb{Z}^d} \right\|_{\ell^\tau} \le C \|f\|_{M^\tau}.$$

As $\tau < p$ then $M^\tau \subseteq M^p$ (cf. [20]), which shows that $f \in \mathcal{K}_\tau^\tau(\mathcal{D}, M^p)$ and $|f|_{\mathcal{K}_\tau^\tau(\mathcal{D},M^p)} \leq C\|f\|_{M^\tau}$. The converse embedding follows from

$$\|f\|_{M^\tau} = \|R\boldsymbol{c}\|_{M^\tau} \leq C' \, \|\boldsymbol{c}\|_{\ell^\tau} = C'|f|_{\mathcal{K}_\tau^\tau(\mathcal{D},M^p)}, \quad f \in \mathcal{K}_\tau^\tau(\mathcal{D}, M^p).$$

This completes the proof. □

In the next section we apply the proposed method for denoising music signals with elements from a Gabor dictionary. For an introduction to Gabor theory in the finite settings, we refer the reader to [35, 37].

5 Numerical experiments

For the implementation we use MATLAB 2017B and apply the routines from the following two toolboxes: The "Large time-frequency analysis toolbox" (LTFAT) version 2.2.0 [31] avaliable from `http://ltfat.sourceforge.net/` and the StrucAudioToolbox [33] available from `http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html`. All Gabor transforms are constructed using 1024 frequency channels, a hop size of 256, and a Hanning window of length 1024 (this is the default settings in the StrucAudioToolbox). These settings lead to transforms of redundancy of four, meaning there are four times as many time-frequency coefficients as signal samples. The music signals we consider are part of the EBU-SQAM database [39], which consists of 70 test sounds sampled at 44 kHz. The database contains a large variety of different music sounds including single instruments, vocal, and orchestra. We measure the reconstruction error of an algorithm by the relative root mean square (RMS) error

$$\mathrm{RMS}(f, f_{rec}) := \frac{\|f - f_{rec}\|_2}{\|f\|_2}.$$

We begin by analyzing the first 524288 samples of signal 8 in the EBU-SQAM database, which consists of an increasing melody of 10 tones played by a violin. A noisy version of the signal is constructed by adding white Gaussian noise and the resulting spectrograms can be found in Fig. A.1.

For the task of denoising we first compare the greedy thresholding approach from nonlinear approximation theory (cf. (A.6)) with the Windowed-Group-Lasso (WGL) from social sparsity. For the WGL we use the default settings of the StrucAudioToolbox, which applies a horizontal asymmetric neighborhood for the shrinkage operator, see [33] for further details. The WGL constructs a denoised version of the spectrogram using only 74739 non-zero coefficients (out of a total of 1050624 coefficients). Using the same number of non-zero coefficients for the greedy thresholding approach we obtain the spectrograms shown in Fig. A.2.

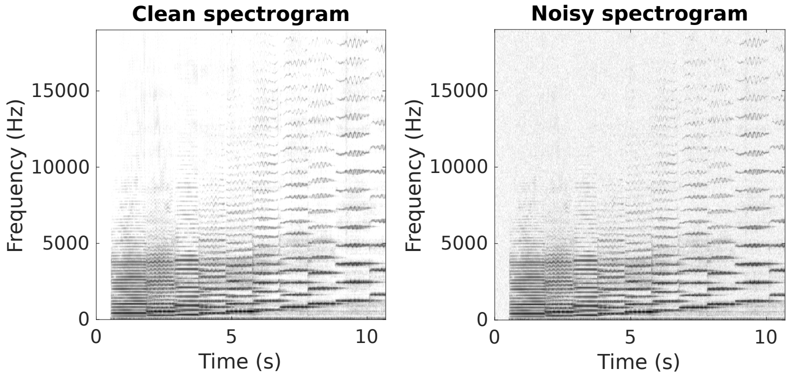

Fig. A.1: Clean and noisy spectrograms of violin music.

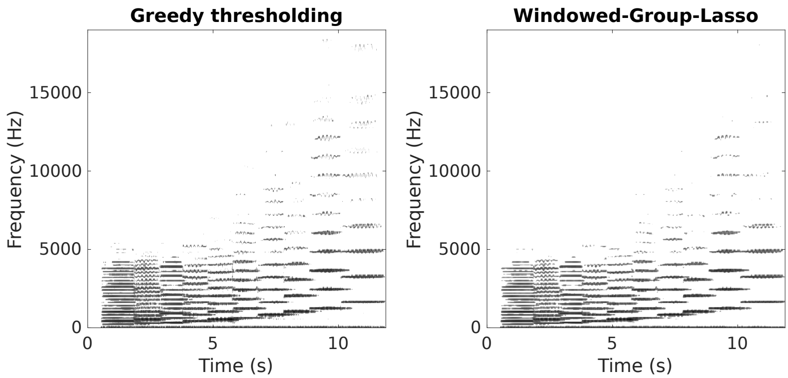

Fig. A.2: Greedy thresholding and the WGL with 74739 non-zero coefficients.

We note from Fig. A.2 that the greedy thresholding approach includes more coefficients at higher frequencies than the WGL. On the other hand, the WGL includes more coefficients at lower frequencies, resulting in a smoother resolution for the fundamental frequencies and the first harmonics. This illustrates the way the WGL is designed, namely that a large isolated coefficient may be discarded in exchange for a smaller coefficient with large neighbors. The RMS error is ≈ 0.084 for the WGL and ≈ 0.034 for the greedy thresholding algorithm. To visualize the performance of the proposed algorithm we choose a rather extreme (horizontal) weight with

$$c^{\Lambda}_{m,n} = |c_{m,n-2}| + |c_{m,n-1}| + |c_{m,n}| + |c_{m,n+1}| + |c_{m,n+2}| \,. \tag{A.12}$$

We then choose the 74739 coefficients with largest weighted magnitudes according to (A.12). In Fig. A.3 we have compared the resulting spectrogram against the spectrogram obtained using the greedy thresholding approach.

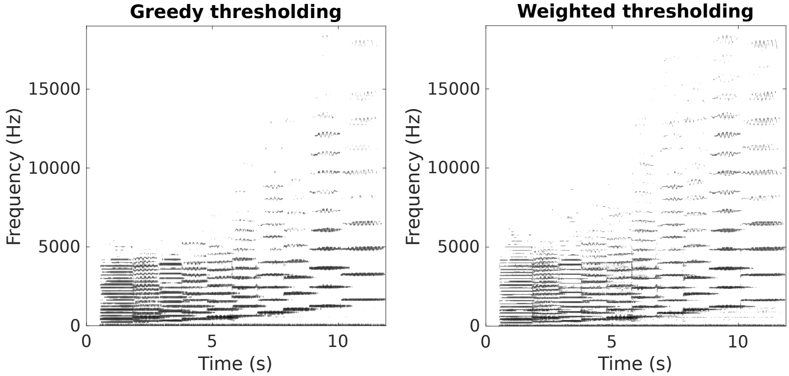

Fig. A.3: Greedy thresholding and weighted thresholding with 74739 non-zero coefficients.

We can see from Fig. A.3 that the horizontal weight in (A.12) enforces the structures at higher frequencies even further than the greedy thresholding approach. This might be desirable for some applications as the timbre of the instrument is determined by the harmonics. The RMS error associated with the weighted thresholding approach is ≈ 0.074, which is higher than for greedy thresholding but lower than the WGL. Applying a more moderate weight (for instance Weight 2 defined in (A.13) below) we obtain an RMS error of ≈ 0.031, which is lower than for greedy thresholding.

We now extend the experiment described above to the entire EBU-SQAM database. For the weighted thresholding we consider the following three weights

$$\begin{aligned} &\text{Weight 1: } c^{\Lambda}_{m,n} = |c_{m,n}| + (|c_{m-1,n}| + |c_{m+1,n}|)/2, \\ &\text{Weight 2: } c^{\Lambda}_{m,n} = |c_{m,n}| + (|c_{m,n-1}| + |c_{m,n+1}|)/2, \\ &\text{Weight 3: } c^{\Lambda}_{m,n} = |c_{m,n}| + (|c_{m,n-1}| + |c_{m-1,n}| + |c_{m,n+1}| + |c_{m+1,n}|)/4. \end{aligned} \tag{A.13}$$

For each of the 70 test sounds in the EBU-SQAM database, we apply the WGL and calculate the associated RMS reconstruction error and number of non-zero coefficients. Using the same number of non-zero coefficients, we then apply the greedy thresholding approach and the three weighted thresholding approaches defined in (A.13). The resulting averaged values can be found in in Table A.1.

Table A.1: Average RMS errors over the EBU-SQAM database for the WGL, the greedy thresholding approach, and the three weighted thresholding approaches defined in (A.13).

Algorithm:	WGL	Greedy	Weight 1	Weight 2	Weight 3
Average RMS error.:	0.1031	0.0462	0.0511	0.0453	0.0472

The average number of coefficients was 68983, which corresponds to 6.5% of the total number of coefficients. We note that the RMS error associated with the WGL is roughly twice as large as for the various thresholding algorithms. We also note that the smallest RMS error is obtained by the weighted thresholding approach, which applies a horizontal weight (Weight 2). This is likely due to the horizontal structure of the harmonics as seen in Fig. A.1.

Let us mention that it might be possible to reduce the error for the WGL by tuning the parameters instead of using the default settings (cf. [34] for a detailed analysis of the parameter settings for the WGL). On the other hand, the same holds true for the thresholding algorithms. In the experiment described above we have used the same number of non-zero coefficients as was chosen by the WGL. This is not likely to be optimal for the thresholding algorithms since the optimal number of non-zero coefficients usually depends on the sparsity of the signal (few coefficients for sparse signals and vice versa). Finally, we have not addressed the resulting audio quality of the denoised sounds. In general, it is very hard to decide which algorithm sounds "the best" as this depends on the application and the subjective opinion of the listener. However, there are indeed audible differences between the algorithms. For the violin music in Fig. A.1, the WGL does the best job of removing the noise, but at the price of a poor timbre of the resulting sound. As we include more coefficients at the higher frequencies, the original timbre of the instrument improves together with an increase in noise.

6 Conclusion

We have presented a new thresholding algorithm and proven an associated strong Jackson embedding under rather general conditions. The algorithm extends the classical greedy approach by incorporating a weight function, which exploits the structure of the expansion coefficients. In particular, the algorithm applies to approximation in modulation spaces using Gabor frames. As an application we have considered the task of denoising music signals and compared the proposed method with the greedy thresholding approach and the WGL from social sparsity. The numerical experiments show that the proposed method can be used both for improving the time-frequency resolution and for reducing the RMS error compared to the other two algorithms. The experiments also show that the performance of the algorithm depends crucially on the choice of weight function, which should be adapted to the particular signal class under consideration.

References

[1] L. Borup and M. Nielsen. Frame decomposition of decomposition spaces. *J. Fourier Anal. Appl.*, 13(1):39–70, 2007.

[2] P. L. Butzer and K. Scherer. Jackson and Bernstein-type inequalities for families of commutative operators in Banach spaces. *J. Approximation Theory*, 5:308–342, 1972.

[3] M. a. J. Carro, J. A. Raposo, and J. Soria. Recent developments in the theory of Lorentz spaces and weighted inequalities. *Mem. Amer. Math. Soc.*, 187(877):xii+128, 2007.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, Jan. 2001.

[5] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], second edition, 2016.

[6] C. Darken, M. Donahue, L. Gurvits, and E. Sontag. Rate of approximation results motivated by robust neural network learning. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, COLT '93, pages 303–309, New York, NY, USA, 1993. ACM.

[7] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximation*, 13(1):57–98, Mar 1997.

[8] R. A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[9] R. A. DeVore. Nonlinear approximation and its applications. In *Multiscale, nonlinear and adaptive approximation*, pages 169–201. Springer, Berlin, 2009.

[10] R. A. DeVore, B. Jawerth, and V. Popov. Compression of wavelet decompositions. *Amer. J. Math.*, 114(4):737–785, 1992.

[11] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[12] R. A. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Advances in Computational Mathematics*, 5(1):173–187, Dec 1996.

[13] H. G. Feichtinger. Modulation spaces on locally compact abelian groups. Technical report, University of Vienna, 1983.

[14] H. G. Feichtinger. Banach spaces of distributions defined by decomposition methods. II. *Math. Nachr.*, 132:207–237, 1987.

[15] H. G. Feichtinger and P. Gröbner. Banach spaces of distributions defined by decomposition methods. I. *Math. Nachr.*, 123:97–120, 1985.

[16] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. I. *J. Funct. Anal.*, 86(2):307–340, 1989.

[17] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. II. *Monatsh. Math.*, 108(2-3):129–148, 1989.

[18] R. Gribonval and M. Nielsen. Some remarks on non-linear approximation with Schauder bases. *East J. Approx.*, 7(3):267–285, 2001.

[19] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51–71, 2004.

[20] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[21] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–331, 2000.

[22] R. S. Ismagilov. Diameters of sets in normed linear spaces, and the approximation of functions by trigonometric polynomials. *Uspehi Mat. Nauk*, 29(3(177)):161–178, 1974.

[23] M. Kowalski. Sparse regression using mixed norms. *Appl. Comput. Harmon. Anal.*, 27(3):303–324, 2009.

[24] M. Kowalski, K. Siedenburg, and M. Dörfler. Social sparsity! neighborhood systems enrich structured shrinkage operators. *IEEE Transactions on Signal Processing*, 61(10):2498–2511, May 2013.

[25] M. Kowalski and B. Torrésani. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. *Signal, Image and Video Processing*, 3(3):251–264, Sep 2009.

[26] M. Kowalski and B. Torrésani. Structured sparsity: from mixed norms to structured shrinkage. In R. Gribonval, editor, *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, Saint Malo, France, Apr. 2009. Inria Rennes - Bretagne Atlantique.

[27] K. I. Oskolkov. Polygonal approximation of functions of two variables. *Mat. Sb. (N.S.)*, 107(149)(4):601–612, 639, 1978.

[28] E. S. Ottosen and M. Nielsen. A characterization of sparse nonstationary Gabor expansions. *Journal of Fourier Analysis and Applications*, May 2017.

[29] P. P. Petrushev. Direct and converse theorems for spline and rational approximation and Besov spaces. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 363–377. Springer, Berlin, 1988.

[30] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies. Sparse representations in audio and music: From coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005, June 2010.

[31] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs. The Large Time-Frequency Analysis Toolbox 2.0. In M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, editors, *Sound, Music, and Motion*, Lecture Notes in Computer Science, pages 419–442. Springer International Publishing, 2014.

[32] E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Math. Ann.*, 63(4):433–476, 1907.

[33] K. Siedenburg and M. Dörfler. Structured sparsity for audio signals. In *Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11)*, Paris, France, 2011.

[34] K. Siedenburg and M. Dörfler. Audio denoising by generalized time-frequency thresholding. In *Proceedings of the AES 45th Conference on Applications of Time-Frequency Processing*, Helsinki, Finland, 2012.

[35] P. L. Søndergaard. *Finite discrete Gabor analysis*. PhD thesis, Technical University of Denmark, 2007.

[36] S. B. Stechkin. On absolute convergence of orthogonal series. *Dokl. Akad. Nauk SSSR (N.S.)*, 102:37–40, 1955.

[37] T. Strohmer. Numerical algorithms for discrete Gabor expansions. In *Gabor analysis and algorithms*, Appl. Numer. Harmon. Anal., pages 267–294. Birkhäuser Boston, Boston, MA, 1998.

[38] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288, 1996.

[39] E. B. Union. Sound quality assessment material: Recordings for subjective tests ; User's handbook for the EBU - SQAM compact disc, Sep 2008.

[40] F. Voigtlaender and A. Pein. Analysis vs. synthesis sparsity for α-shearlets. *ArXiv preprint*, arXiv:1702.03559v1, 2017.

Paper B

A Characterization of Sparse Nonstationary Gabor Expansions

Emil Solsbæk Ottosen and Morten Nielsen

The paper has been accepted for publication in
Journal of Fourier Analysis and Applications, 2017.

Abstract

We investigate the problem of constructing sparse time-frequency representations with flexible frequency resolution, studying the theory of nonstationary Gabor frames in the framework of decomposition spaces. Given a painless nonstationary Gabor frame, we construct a compatible decomposition space and prove that the nonstationary Gabor frame forms a Banach frame for the decomposition space. Furthermore, we show that the decomposition space norm can be completely characterized by a sparseness condition on the frame coefficients and we prove an upper bound on the approximation error occurring when thresholding the frame coefficients for signals belonging to the decomposition space.

1 Introduction

Redundant Gabor frames play an essential role in time-frequency (TF) analysis as these frames provide expansions with good TF resolution [6, 27]. Gabor frames are based on translation and modulation of a single window function according to lattice parameters which largely determine the redundancy of the frame. By varying the support of the window function one can change the overall resolution of the frame, but it is in general not possible to change the resolution in specific regions of the TF plane. For signals with varying TF characteristics, a fixed resolution is often undesirable. To overcome this problem, the usage of *multi-window Gabor frames* has been proposed [10, 34, 42, 43]. As opposed to standard Gabor frames, multi-window Gabor frames use a whole catalogue of window functions of different shapes and sizes to create adaptive representations. A recent example is the *nonstationary Gabor frames* (NSGFs) which have shown great potential in capturing the essential TF information of music signals [1, 11, 32, 33]. These frames use different window functions along *either* the time- or the frequency axes and guarantee perfect reconstruction and an FFT-based implementation in the *painless case*. Originally, NSGFs were studied by Hernández, Labate & Weiss [30] and later by Ron & Shen [40] who named them *generalized shift-invariant systems*. We choose to work with the terminology introduced in [1] as we will only consider frames in the painless case for which several practical implementations have been constructed under the name of NSGFs [1, 11, 32]. We consider painless NSGFs with flexible frequency resolution, corresponding to a sampling grid in the TF plane which is irregular over frequency but regular over time at each fixed frequency position. This construction is particularly useful in connection with music signals since the NSGF can be set to coincide with the semitones used in Western music. Based on the nature of musical tones [9, 38], we expect music signals to permit sparse expansions relative to the redundant NSGF dictionaries.

The main contribution of this paper is a theoretical characterization of the signals with sparse expansions relative to the NSGF dictionaries. By a sparse expansion we mean an expansion for which the original signal can be approximated at a certain rate by thresholding the expansion coefficients. To prove such a characterization, we follow the approach in [22, 23, 28] and search for a smoothness space compatible with the structure of the frame. Classical smoothness spaces such as modulation spaces [15] or Besov spaces [41] cannot be expected to be linked with sparse expansions relative to the NSGF dictionaries since these smoothness spaces are not compatible with the flexible frequency resolution of the NSGFs. Modulation spaces correspond to a uniform partition of the frequency domain while Besov spaces correspond to a dyadic partition. Therefore, we study NSGFs in the framework of decomposition spaces. Decomposition spaces were introduced by Feichtinger & Gröbner in [17], and further studied by Feichtinger in [16], and form a large class of function spaces on $\mathbb{R}^d$ including smoothness spaces such as modulation spaces, Besov spaces, and the intermediate $\alpha-$modulation spaces as special cases [2, 3, 25]. We construct the decomposition spaces using structured coverings, as introduced by Borup & Nielsen in [3], which leads to a partition of the frequency domain obtained by applying invertible affine transformations $\{A_k(\cdot) + c_k\}_{k\in\mathbb{N}}$ on a fixed set $Q \subset \mathbb{R}^d$.

Given a painless NSGF, we provide a method for constructing a compatible structured covering and the associated decomposition space. We then show that the NSGF forms a Banach frame for the decomposition space and prove that signals belong to the decomposition space if and only if they permit sparse frame expansions. Based on the sparse expansions, we prove an upper bound on the approximation error occurring when thresholding the frame coefficients for signals belonging to the decomposition space. All these results are based on the characterization given in Theorem 5.1 which is the main contribution of this article. This theorem yields the existence of constants $0 < C_1, C_2 < \infty$ such that all signals f, belonging to the decomposition space $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$, satisfy

$$C_1 \|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \leq \left\| \left\{ \left\langle f, h^p_{T,n} \right\rangle \right\}_{T,n} \right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})} \leq C_2 \|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})},$$

with $\{h^p_{T,n}\}_{T,n}$ denoting L^p-normalized elements from the NSGF and $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ an associated sequence space. In this way we completely characterize the decomposition space using the frame coefficients from the NSGF.

The outline of the article is as follows. In Section 2 we define decomposition spaces based on structured coverings and in Section 3 we define NSGFs in the notation of [1]. We construct the compatible decomposition space in Section 4 and in Section 5 we prove Theorem 5.1. In Section 6 we show that the NSGF forms a Banach frame for the compatible decomposition space and in Section 7 we provide the link to nonlinear approximation theory.

Let us now introduce some of the notation used throughout this article. We let $\hat{f}(\xi) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i x\cdot\xi}dx$ denote the Fourier transform with the usual extension to $L^2(\mathbb{R}^d)$. By $F \asymp G$ we mean that there exist two constants $0 < C_1, C_2 < \infty$ such that $C_1F \leq G \leq C_2F$. For two (quasi-)normed vector spaces X and Y, $X \hookrightarrow Y$ means that $X \subset Y$ and $\|f\|_Y \leq C\|f\|_X$ for some constant C and all $f \in X$. We say that a non-empty open set $\Omega' \subset \mathbb{R}^d$ is *compactly contained* in an open set $\Omega \subset \mathbb{R}^d$ if $\overline{\Omega'} \subset \Omega$ and $\overline{\Omega'}$ is compact. We denote the matrix norm $\max\{|a_{ij}|\}$ by $\|A\|_{\ell^\infty(\mathbb{R}^{d\times d})}$ and we call $\{\xi_i\}_{i\in\mathcal{I}} \subset \mathbb{R}^d$ a $\delta-$separated set if $\inf_{j,k\in\mathcal{I},j\neq k}\|\xi_j - \xi_k\|_2 = \delta > 0$. Finally, by I_d we denote the identity operator on $\mathbb{R}^d$ and by χ_Q we denote the indicator function for a set $Q \subset \mathbb{R}^d$.

2 Decomposition spaces

In order to construct decomposition spaces, we first need the notion of a structured covering with an associated bounded admissible partitions of unity (BAPU) as defined in Section 2.1. A BAPU defines a (flexible) partition of the frequency domain corresponding to the structured covering. We use the notation of [3] but with slightly modified definitions for both the structured coverings and the BAPUs.

2.1 Structured covering and BAPU

For an invertible matrix $A \in GL(\mathbb{R}^d)$, and a constant $c \in \mathbb{R}^d$, we define the affine transformation

$$T\xi := A\xi + c, \quad \xi \in \mathbb{R}^d.$$

For a subset $Q \subset \mathbb{R}^d$ we let $Q_T := T(Q)$, and for notational convenience we define $|T| := |\det(A)|$. Given a family $\mathcal{T} = \{A_k(\cdot) + c_k\}_{k\in\mathbb{N}}$ of invertible affine transformations on $\mathbb{R}^d$, and a subset $Q \subset \mathbb{R}^d$, we set $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$ and

$$\widetilde{T} := \left\{T' \in \mathcal{T} \mid Q_{T'} \cap Q_T \neq \emptyset\right\}, \quad T \in \mathcal{T}. \tag{B.1}$$

We say that $\mathcal{Q}$ is an *admissible covering* of $\mathbb{R}^d$ if $\bigcup_{T\in\mathcal{T}} Q_T = \mathbb{R}^d$ and there exists $n_0 \in \mathbb{N}$ such that $|\widetilde{T}| \leq n_0$ for all $T \in \mathcal{T}$. We note that the (minimal) number n_0 is the degree of overlap between the sets constituting the covering.

Definition 2.1 ($\mathcal{Q}-$moderate weight)**.** Let $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$ be an admissible covering. A function $u : \mathbb{R}^d \to (0,\infty)$ is called $\mathcal{Q}-$moderate if there exists $C > 0$ such that $u(x) \leq Cu(y)$ for all $x, y \in Q_T$ and all $T \in \mathcal{T}$. A $\mathcal{Q}-$moderate weight (derived from u) is a sequence $\{\omega_T\}_{T\in\mathcal{T}} := \{u(\xi_T)\}_{T\in\mathcal{T}}$ with $\xi_T \in Q_T$ for all $T \in \mathcal{T}$.

For the rest of this article we shall use the explicit choice $u(\xi) := 1 + \|\xi\|_2$ for the function u in Definition 2.1. We now define the concept of a structured covering, first considered in [3]. To ensure that the resulting decomposition spaces are complete, we consider an extended version of the definition given in [3].

Definition 2.2 (Structured covering)**.** Given a family $\mathcal{T} = \{A_k(\cdot) + c_k\}_{k\in\mathbb{N}}$ of invertible affine transformations on $\mathbb{R}^d$, suppose there exist two bounded open sets $P \subset Q \subset \mathbb{R}^d$, with P compactly contained in Q, such that

1. $\{P_T\}_{T\in\mathcal{T}}$ and $\{Q_T\}_{T\in\mathcal{T}}$ are admissible coverings.
2. There exists $K > 0$, such that $\left\|A_{k'}^{-1}A_k\right\|_{\ell^\infty(\mathbb{R}^{d\times d})} \leq K$ holds whenever $(A_{k'}Q + c_{k'}) \cap (A_kQ + c_k) \neq \emptyset$.
3. There exists $K_* > 0$, such that $\left\|A_k^{-1}\right\|_{\ell^\infty(\mathbb{R}^{d\times d})} \leq K_*$ holds for all $k \in \mathbb{N}$.
4. There exists a $\delta-$separated set $\{\xi_T\}_{T\in\mathcal{T}} \subset \mathbb{R}^d$, with $\xi_T \in Q_T$ for all $T \in \mathcal{T}$, such that $\{\omega_T\}_{T\in\mathcal{T}} := \{u(\xi_T)\}_{T\in\mathcal{T}}$ is a $\mathcal{Q}-$moderate weight.
5. There exists $\gamma > 0$, such that $|Q_T| \leq \omega_T^\gamma$ for all $T \in \mathcal{T}$.

Then we call $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ a *structured covering*.

Remark 2.1. Definition 2.2(3)-(5) are new additions compared to the definition given in [3] and are necessary for proving Theorem 2.1 page 65. We note that Definition 2.2(2) implies $|Q_{T'}| \asymp |Q_T|$ uniformly for all $T \in \mathcal{T}$ and all $T' \in \widetilde{T}$, and Definition 2.2(3) implies a uniform lower bound on $|Q_T|$.

For a structured covering we have the associated concept of a BAPU, first considered in [3, 17]. With a small modification of the proof of [3, Proposition 1] we have the following result.

Proposition 2.1. *Given a structured covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$, there exists a family of non-negative functions $\{\psi_T\}_{T\in\mathcal{T}} \subset C_c^\infty(\mathbb{R}^d)$ satisfying*

1. $\operatorname{supp}(\psi_T) \subset Q_T$ *for all* $T \in \mathcal{T}$.
2. $\sum_{T\in\mathcal{T}} \psi_T(\xi) = 1$ *for all* $\xi \in \mathbb{R}^d$.
3. $\sup_{T\in\mathcal{T}} |Q_T|^{1/p-1} \left\|\mathcal{F}^{-1}\psi_T\right\|_{L^p} < \infty$ *for all* $0 < p \leq 1$.
4. *For all $\alpha \in \mathbb{N}_0^d$, there exists $C_\alpha > 0$ such that $|\partial^\alpha \psi_T(\xi)| \leq C_\alpha \chi_{Q_T}(\xi)$, for all $\xi \in \mathbb{R}^d$ and all $T \in \mathcal{T}$.*

We say that $\{\psi_T\}_{T\in\mathcal{T}}$ is a BAPU subordinate to $\mathcal{Q}$.

Remark 2.2. Proposition 2.1(3) is necessary to ensure that the decomposition spaces under consideration will be well-defined for $0 < p < 1$. This case is of specific interest since it plays an essential role in connection with nonlinear approximation theory (cf. Section 7).

Remark 2.3. Proposition 2.1(4) is a new addition compared to [3, Proposition 1] and is necessary for proving Theorem 2.1 page 65. The proof of Proposition 2.1(4) follows easily from the arguments in the proof of [3, Proposition 1] and Definition 2.2(3). Finally, it should be noted that the assumptions in Definition 2.2(4)-(5) are not necessary for proving Proposition 2.1, however, these assumptions are needed for the proof of Theorem 2.1.

The proof of [3, Proposition 1] is constructive and provides a method for constructing the associated BAPU. Given a structured covering $\{Q_T\}_{T\in\mathcal{T}}$ (with P being compactly contained in Q), the method goes as follows:

1. Pick a non-negative function $\Phi \in C_c^\infty(\mathbb{R}^d)$ with $\Phi(\xi) = 1$ for all $\xi \in P$ and $\operatorname{supp}(\Phi) \subset Q$.
2. For all $T \in \mathcal{T}$, define
$$\psi_T(\xi) = \frac{\Phi(T^{-1}\xi)}{\sum_{T'\in\mathcal{T}} \Phi(T'^{-1}\xi)}.$$
3. Then $\{\psi_T\}_{T\in\mathcal{T}}$ is a BAPU subordinate to $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$.

In the next section we define decomposition spaces based on structured coverings.

2.2 Definition of decomposition spaces

Given a structured covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ with corresponding $\mathcal{Q}-$moderate weight $\{\omega_T\}_{T\in\mathcal{T}}$ and BAPU $\{\psi_T\}_{T\in\mathcal{T}}$. For $s \in \mathbb{R}$ and $0 < q \leq \infty$, we define the associated weighted sequence space $\ell^q_{\omega^s}(\mathcal{T})$ as the sequences of complex numbers $\{a_T\}_{T\in\mathcal{T}}$ satisfying

$$\|\{a_T\}_{T\in\mathcal{T}}\|_{\ell^q_{\omega^s}} := \|\{\omega_T^s a_T\}_{T\in\mathcal{T}}\|_{\ell^q} < \infty.$$

Given $\{a_T\}_{T\in\mathcal{T}} \in \ell^q_{\omega^s}(\mathcal{T})$, we define $\{a_T^+\}_{T\in\mathcal{T}}$ by $a_T^+ := \sum_{T'\in\widetilde{T}} a_{T'}$. Since $\{\omega_T\}_{T\in\mathcal{T}}$ is $\mathcal{Q}-$moderate, $\{a_T\}_{T\in\mathcal{T}} \to \{a_T^+\}_{T\in\mathcal{T}}$ defines a bounded operator on $\ell^q_{\omega^s}(\mathcal{T})$ [17, Remark 2.13 and Lemma 3.2]. Denoting its operator norm by C_+, we have

$$\left\|\{a_T^+\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}} \leq C_+ \left\|\{a_T\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}}, \quad \forall \{a_T\}_{T\in\mathcal{T}} \in \ell^q_{\omega^s}(\mathcal{T}). \tag{B.2}$$

We will use (B.2) several times throughout this article. Using the notation of [3] we define the Fourier multiplier $\psi_T(D)$ by

$$\psi_T(D)f := \mathcal{F}^{-1}(\psi_T \mathcal{F} f), \quad f \in L^2(\mathbb{R}^d).$$

Combining Proposition 2.1(3) with Lemma A.2 page 81 and [3, Lemma 1] we can show the existence of a uniform constant $C > 0$ such that all band-limited functions $f \in L^p(\mathbb{R}^d)$ satisfy

$$\|\psi_T(D)f\|_{L^p} \leq C \|f\|_{L^p},$$

for all $T \in \mathcal{T}$ and all $0 < p \leq \infty$. That is, $\psi_T(D)$ extends to a bounded operator on the band-limited functions in $L^p(\mathbb{R}^d)$, uniformly in $T \in \mathcal{T}$. Let us now give the definition of decomposition spaces on the Fourier side.

Definition 2.3 (Decomposition space)**.** Let $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ be a structured covering of $\mathbb{R}^d$ with corresponding $\mathcal{Q}-$moderate weight $\{\omega_T\}_{T\in\mathcal{T}}$ and subordinate BAPU $\{\psi_T\}_{T\in\mathcal{T}}$. For $s \in \mathbb{R}$ and $0 < p, q < \infty$, we define the *decomposition space* $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ as the set of distributions $f \in \mathcal{S}'(\mathbb{R}^d)$ satisfying

$$\|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} := \left\| \{\|\psi_T(D)f\|_{L^p}\}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}} < \infty.$$

Remark 2.4. According to [17, Theorem 3.7], two different BAPUs yield the same decomposition space with equivalent norms so $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ is in fact well defined and independent of the choice of BAPU. Actually, the results in [17] show that decomposition spaces are invariant under certain geometric modifications of the covering $\mathcal{Q}$, but we will not go into detail here.

Remark 2.5. In their most general form, decomposition spaces $D(\mathcal{Q}, B, Y)$ are constructed using a *local component* B and a *global component* Y [17]. This construction is similar to the construction of Wiener amalgam spaces $W(B, C)$ with local component B and global component C [14, 29, 39]. However, Wiener amalgam spaces are based on bounded *uniform* partitions of unity, which corresponds to a uniform upper bound on the size of the members of the covering. We do not find such an assumption natural in relation to NSGFs (cf. Section 3) and have therefore chosen the more general framework of decomposition spaces.

In Theorem 2.1 below we prove that $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ is in fact a (quasi-)Banach space. Before presenting this result, let us first consider some examples of familiar decomposition spaces. By standard arguments, one can easily show that $D(\mathcal{Q}, L^2, \ell^2) = L^2(\mathbb{R}^d)$ with equivalent norms for any structured covering $\mathcal{Q}$. The next two examples are not as straightforward and demand some structure on the covering. Recall that $\{\xi_T\}_{T\in\mathcal{T}}$ denotes the $\delta-$separated set from Definition 2.2(4).

Example 2.1 (Modulation spaces). Let $Q \subset \mathbb{R}^d$ be an open cube with center 0 and side length $r > 1$. Define $\mathcal{T} := \{T_k\}_{k\in\mathbb{Z}^d}$, with $T_k\xi := \xi - k$, and set $\xi_{T_k} := k$ for all $k \in \mathbb{Z}^d$. With $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$ then $D(\mathcal{Q}, L^p, \ell^q_{\omega^s}) = M^s_{p,q}(\mathbb{R}^d)$ for $s \in \mathbb{R}$ and $0 < p, q < \infty$, see [15, Section 4] for further details. △

Example 2.2 (Besov spaces). Let $E_2 := \{\pm 1, \pm 2\}$, $E_1 := \{\pm 1\}$ and $E := E_2^d \setminus E_1^d$. For $j \in \mathbb{N}$ and $k \in E$ define $c_{j,k} := 2^j(v(k_1), \ldots, v(k_d))$, where

$$v(x) := \operatorname{sgn}(x) \cdot \begin{cases} 1/2 & \text{for } x = \pm 1 \\ 3/2 & \text{for } x = \pm 2 \end{cases}$$

Let $Q \subset \mathbb{R}^d$ be an open cube with center 0 and side length $r > 2$. Define $\mathcal{T} := \{I, T_{j,k}\}_{j\in\mathbb{N},k\in E}$, with $T_{j,k}\xi := 2^j\xi + c_{j,k}$, and set $\xi_{T_{j,k}} := c_{j,k}$ for all $j \in \mathbb{N}$ and $k \in E$. With $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$ then $D(\mathcal{Q}, L^p, \ell^q_{\omega^s}) = B^s_{p,q}(\mathbb{R}^d)$ for $s \in \mathbb{R}$ and $0 < p, q < \infty$, see [41, Section 2.5.4] for further details. △

Let us now study some important properties of decomposition spaces, in particular completeness.

Theorem 2.1. *Given a structured covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ with $\mathcal{Q}-$moderate weight $\{\omega_T\}_{T\in\mathcal{T}}$ and subordinate BAPU $\{\psi_T\}_{T\in\mathcal{T}}$. For $s \in \mathbb{R}$ and $0 < p, q < \infty$,*

1. $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D(\mathcal{Q}, L^p, \ell^q_{\omega^s}) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$.
2. $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ *is a (quasi-)Banach space (Banach space if* $1 \le p, q < \infty$*).*
3. $\mathcal{S}(\mathbb{R}^d)$ *is dense in* $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$.

Remark 2.6. As was pointed out in [21], the definition of decomposition spaces given in [3] cannot guarantee completeness in the general case. However in [4], this problem was fixed by imposing certain weight conditions on the structured covering. Our proof of Theorem 2.1 is based on the approach taken in [4].

In Appendix A we have provided a sketch of the proof for Theorem 2.1. The underlying ideas for the proof are similar to those of [4, Proposition 5.2] and several references are made to results in [4]. However, in [4] the authors considered only coverings made up from open balls and not all arguments carry over to the general case of an arbitrary structured covering.

3 Nonstationary Gabor frames

In this section we define nonstationary Gabor frames with flexible frequency resolution using the notation of [1]. Given a set of window functions $\{h_m\}_{m\in\mathbb{Z}^d}$ in $L^2(\mathbb{R}^d)$, with corresponding time sampling steps $a_m > 0$, for $m, n \in \mathbb{Z}^d$ we define atoms of the form

$$h_{m,n}(x) := h_m(x - na_m), \quad x \in \mathbb{R}^d.$$

The choice of $\mathbb{Z}^d$ as index set for m is only a matter of notational convenience; any countable index set would do. If $\sum_{m,n} |\langle f, h_{m,n}\rangle|^2 \asymp \|f\|_2^2$ for all $f \in L^2(\mathbb{R}^d)$, we refer to $\{h_{m,n}\}_{m,n}$ as a *nonstationary Gabor frame* (NSGF). For an NSGF $\{h_{m,n}\}_{m,n}$, the frame operator

$$Sf = \sum_{m,n\in\mathbb{Z}^d} \langle f, h_{m,n}\rangle\, h_{m,n}, \quad f \in L^2(\mathbb{R}^d),$$

is invertible and we have the expansions

$$f = \sum_{m,n\in\mathbb{Z}^d} \langle f, h_{m,n}\rangle\, \tilde{h}_{m,n}, \quad f \in L^2(\mathbb{R}^d),$$

with $\{\tilde{h}_{m,n}\}_{m,n} := \{S^{-1}h_{m,n}\}_{m,n}$ being the canonical dual frame of $\{h_{m,n}\}_{m,n}$. An NSGF with flexible frequency resolution corresponds to a grid in the time-frequency plane which is irregular over frequency but regular over time at each frequency position. This property allows for adaptive time-frequency representations as opposed to standard Gabor frames. According to [1, Corollary 2], we have the following important result for NSGFs with band-limited window functions.

Theorem 3.1. *Let $\{h_m\}_{m\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ with time sampling steps $\{a_m\}_{m\in\mathbb{Z}^d}$, $a_m > 0$ for all $m \in \mathbb{Z}^d$. Assuming* $\operatorname{supp}(\hat{h}_m) \subseteq [0, \frac{1}{a_m}]^d + b_m$, *with $b_m \in \mathbb{R}^d$ for all $m \in \mathbb{Z}^d$, the frame operator for the system*

$$h_{m,n}(x) = h_m(x - na_m), \quad \forall m, n \in \mathbb{Z}^d, \quad x \in \mathbb{R}^d,$$

is given by

$$Sf(x) = \left(\mathcal{F}^{-1}\left(\sum_{m\in\mathbb{Z}^d} \frac{1}{a_m^d}\left|\hat{h}_m\right|^2\right) * f\right)(x), \quad f \in L^2(\mathbb{R}^d).$$

The system $\{h_{m,n}\}_{m,n\in\mathbb{Z}^d}$ constitutes a frame for $L^2(\mathbb{R}^d)$, with frame-bounds $0 < A \leq B < \infty$, if and only if

$$A \leq \sum_{m\in\mathbb{Z}^d} \frac{1}{a_m^d}\left|\hat{h}_m(\xi)\right|^2 \leq B, \quad \text{for a.e. } \xi \in \mathbb{R}^d, \tag{B.3}$$

and the canonical dual frame is then given by

$$\tilde{h}_{m,n}(x) = \mathcal{F}^{-1}\left(\frac{\hat{h}_m}{\sum_{l\in\mathbb{Z}^d} \frac{1}{a_l^d}\left|\hat{h}_l\right|^2}\right)(x - na_m), \quad x \in \mathbb{R}^d. \tag{B.4}$$

Remark 3.1. We note that the canonical dual frame in (B.4) posses the same structure as the original frame, which is a property not shared by general NSGFs. We also note that the canonical tight frame can be obtained by taking the square root of the denominator in (B.4).

Traditionally, an NSGF satisfying the assumptions of Theorem 3.1 is called a *painless* NSGF, referring to the fact that the frame operator is simply a multiplication operator (in the frequency domain) and therefore easily invertible. This terminology is adopted from the classical *painless nonorthogonal expansions* [7], which corresponds to the painless case for standard Gabor frames.

By slight abuse of notation we use the term "painless" to denote the NSGFs satisfying Definition 3.1 below. In order to properly formulate this definition, we first need some preliminary notation. Let $\{h_m\}_{m\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ satisfy the assumptions in Theorem 3.1. Given $C_* > 0$ we denote by $\{I_m\}_{m\in\mathbb{Z}^d}$ the open cubes

$$I_m := \left(-\varepsilon_m, \frac{1}{a_m} + \varepsilon_m\right)^d + b_m, \quad m \in \mathbb{Z}^d, \tag{B.5}$$

with $\varepsilon_m := C_*/a_m$ for all $m \in \mathbb{Z}^d$. We note that $\operatorname{supp}(\hat{h}_{m,n}) \subset I_m$ for all $m, n \in \mathbb{Z}^d$. For $m \in \mathbb{Z}^d$ we define

$$\widetilde{m} := \left\{ m' \in \mathbb{Z}^d \mid I_{m'} \cap I_m \neq \emptyset \right\},$$

using the notation of (B.1). With this definition, $|\widetilde{m}|$ denotes the number of cubes overlapping with I_m. Finally, we recall the choice $u(\xi) := 1 + \|\xi\|_2$ for the function u in Definition 2.1.

Definition 3.1 (Painless NSGF)**.** Let $\{h_m\}_{m\in\mathbb{Z}^d} \subset \mathcal{S}(\mathbb{R}^d)$ satisfy the assumptions in Theorem 3.1 and assume that

1. $\{\hat{h}_m\}_{m\in\mathbb{Z}^d} \subset C_c^\infty(\mathbb{R}^d)$ and for $\beta \in \mathbb{N}_0^d$ there exists $C_\beta > 0$, such that
$$\sup_{\xi\in\mathbb{R}^d} \left|\partial_\xi^\beta \hat{h}_m(\xi)\right| \leq C_\beta a_m^{d/2+|\beta|}, \quad \text{for all } m \in \mathbb{Z}^d.$$
2. $\sup_{m\in\mathbb{Z}^d} a_m := a < \infty$.
3. There exists $C_* > 0$ and $n_0 \in \mathbb{N}$, such that the open cubes $\{I_m\}_{m\in\mathbb{Z}^d}$ satisfy $|\widetilde{m}| \leq n_0$ and $a_{m'} \asymp a_m$ uniformly for all $m \in \mathbb{Z}^d$ and all $m' \in \widetilde{m}$.
4. The centerpoints $\{b_m\}_{m\in\mathbb{Z}^d}$ forms a $\delta-$separated set and the sequence $\{\omega_m\}_{m\in\mathbb{Z}^d} := \{u(b_m)\}_{m\in\mathbb{Z}^d}$ constitutes a $\{I_m\}_{m\in\mathbb{Z}^d}-$moderate weight.
5. There exists $\gamma > 0$ such that $|I_m| \leq \omega_m^\gamma$ for all $m \in \mathbb{Z}^d$.

Then we refer to $\{h_{m,n}\}_{m,n\in\mathbb{Z}^d}$ as a *painless* NSGF.

Remark 3.2. Definition 3.1(2) implies a uniform lower bound on $|I_m|$ and Definition 3.1(4) guarantees a minimum distance between the center of the cubes. Furthermore, Definition 3.1(3) implies that each cube I_m has at most n_0 overlap with other cubes and that the side-length of I_m is equivalent to the side-length of any overlapping cube.

Remark 3.3. The assumptions in Definition 3.1 are natural in relation to decomposition spaces and are easily satisfied. However, the support conditions for $\{\hat{h}_m\}_{m\in\mathbb{Z}^d}$ given in Theorem 3.1 are rather restrictive and deserves a discussion. In fact, compact support of the window functions is not a necessary assumption for characterizing modulation spaces [27], Besov spaces [20], or even general decomposition spaces [37]. However, a certain structure of the dual frame is needed and general NSGFs does not provide such structure. We choose to work with the painless case and base our argument on the fact that the dual frame posses the same structure as the original frame. We expect that it is possible to extend the theory developed in this paper to a more general setting by applying existence results for general NSGFs [12, 13, 31] or generalized shift invariant systems [30, 35, 36, 40]. In particular, the paper [31] by Holighaus seems to provide interesting results in this regard. In this paper, it is shown that for compactly supported window functions, the sampling density in Theorem 3.1 can (under mild assumptions) be relaxed such that the dual frame posses a structure similar to that of the original frame. However, it is outside the scope of this paper to include such results and we will not go into further details.

We now provide a simple example of a set of window functions satisfying Definition 3.1(1).

Example 3.1. Let $\varphi \in C_c^\infty(\mathbb{R}^d) \setminus \{0\}$ with $\operatorname{supp}(\varphi) \subseteq [0,1]^d$ and for $m \in \mathbb{Z}^d$ define

$$\hat{h}_m(\xi) := a_m^{d/2}\varphi\left(a_m(\xi - b_m)\right), \quad \forall \xi \in \mathbb{R}^d,$$

with $b_m \in \mathbb{R}^d$ and $a_m > 0$. Then $\operatorname{supp}(\hat{h}_m) \subseteq [0, \frac{1}{a_m}]^d + b_m$. Furthermore, with $w := a_m(\xi - b_m)$ the chain rule yields

$$\left|\partial_\xi^\beta \hat{h}_m(\xi)\right| = \left|\left[\partial_\xi^\beta \varphi\right](w)\right| a_m^{d/2+|\beta|} \leq C_\beta a_m^{d/2+|\beta|} \chi_{[0,\frac{1}{a_m}]^d + b_m}(\xi), \quad \forall \xi \in \mathbb{R}^d.$$

This shows Definition 3.1(1). △

In the next section we consider painless NSGFs in the framework of decomposition spaces in order to characterize signals with sparse expansions relative to the NSGF dictionaries.

4 Decomposition spaces based on nonstationary Gabor frames

We first provide a method for constructing a structured covering which is compatibly with a given painless NSGF $\{h_{m,n}\}_{m,n\in\mathbb{Z}^d} \subset \mathcal{S}(\mathbb{R}^d)$. We recall the definition of $\varepsilon_m = C_*/a_m$ used in the construction of $\{I_m\}_{m\in\mathbb{Z}^d}$ in (B.5). Define $Q := (0,1)^d$ together with the set of affine transformations $\mathcal{T} := \{A_m(\cdot) + c_m\}_{m\in\mathbb{Z}^d}$ with

$$A_m := \left(2\varepsilon_m + \frac{1}{a_m}\right) \cdot I_d \quad \text{and} \quad (c_m)_j := -\varepsilon_m + (b_m)_j, \quad 1 \le j \le d.$$

Then $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}} = \{I_m\}_{m\in\mathbb{Z}^d}$ and, furthermore, we have the following result.

Lemma 4.1. *$\mathcal{Q}$ is a structured covering of $\mathbb{R}^d$.*

Proof. Define the set

$$P := \left(\frac{C_*}{2C_*+1}, \frac{C_*+1}{2C_*+1}\right)^d.$$

By straightforward calculations, it is easy to show that P is compactly contained in Q and $\mathcal{P} := \{P_T\}_{T\in\mathcal{T}} = \{(0, \frac{1}{a_m})^d + b_m\}_{m\in\mathbb{Z}^d}$. Let us now show that $\mathcal{P}$ and $\mathcal{Q}$ satisfy the five conditions of Definition 2.2 page 62.

1. First we show that $\mathcal{P}$ covers $\mathbb{R}^d$. We note that this immediately implies that $\mathcal{Q}$ also covers $\mathbb{R}^d$. Assume $\mathcal{P}$ does not cover $\mathbb{R}^d$, i.e., that there exists some $\xi' \in \mathbb{R}^d$ such that $\xi' \notin (0, \frac{1}{a_m})^d + b_m$ for all $m \in \mathbb{Z}^d$. Since $\operatorname{supp}(\hat{h}_m) \subseteq [0, \frac{1}{a_m}]^d + b_m$, and $\hat{h}_m$ is continuous, we get $\hat{h}_m(\xi') = 0$ for all $m \in \mathbb{Z}^d$. This contradicts the inequality in (B.3) concerning the lower frame bound and thus shows that $\mathcal{P}$ covers $\mathbb{R}^d$. Now, Definition 3.1(3) is precisely the admissibility condition for $\mathcal{Q}$ and thus guarantees that both $\mathcal{P}$ and $\mathcal{Q}$ are admissible coverings. This shows Definition 2.2(1).

2. If $(A_{m'}Q + c_{m'}) \cap (A_m Q + c_m) \neq \emptyset$, then $a_{m'} \asymp a_m$ according to Definition 3.1(3). Furthermore, since $A_{m'}^{-1}A_m$ is a diagonal matrix and $\varepsilon_m = C_*/a_m$ we get

$$\left\|A_{m'}^{-1}A_m\right\|_{\ell^\infty(\mathbb{R}^{d\times d})} = \frac{a_{m'}}{a_m} \le \frac{Ka_m}{a_m} = K,$$

for some $K > 0$, so Definition 2.2(2) is satisfied.

3. To show Definition 2.2(3) we note that

$$\left\|A_m^{-1}\right\|_{\ell^\infty(\mathbb{R}^d\times\mathbb{R}^d)} = \frac{a_m}{2C_*+1} \le \frac{a}{2C_*+1}, \quad \forall m \in \mathbb{Z}^d,$$

according to Definition 3.1(2).

4. Finally, Definition 2.2(4)-(5) follow directly from Definition 3.1(4)-(5).

□

Since $\mathcal{Q}$ is a structured covering, Proposition 2.1 applies and we obtain a BAPU $\{\psi_T\}_{T\in\mathcal{T}}$ subordinate to $\mathcal{Q}$. Given parameters $s \in \mathbb{R}$ and $0 < p, q < \infty$ we may, therefore, construct the associated decomposition space $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$. For notational convenience, we change notation and write $\{h_{T,n}\}_{T\in\mathcal{T},n\in\mathbb{Z}^d}$, such that $\operatorname{supp}(\hat{h}_{T,n}) \subset Q_T$ for all $T \in \mathcal{T}$ and all $n \in \mathbb{Z}^d$. Since $A_T = (2\varepsilon_T + a_T^{-1}) \cdot I_d$, the chain rule and Definition 3.1(1) yield

$$\begin{aligned}\left|\partial_\xi^\beta\left[\hat{h}_T(T\xi)\right]\right| &= \left|\left[\partial_\xi^\beta \hat{h}_T\right](T\xi)\right| \cdot \left(2\varepsilon_T + \frac{1}{a_T}\right)^{|\beta|} \\ &\leq C_\beta a_T^{d/2} \cdot (2C_* + 1)^{|\beta|} \chi_{Q_T}(T\xi) = C'_\beta a_T^{d/2} \chi_Q(\xi), \quad \forall \xi \in \mathbb{R}^d. \end{aligned} \tag{B.6}$$

Using (B.6) we can prove the following decay property of $\{h_{T,n}\}_{T,n}$.

Proposition 4.1. *For every $N \in \mathbb{N}$ there exists a constant $C_N > 0$ such that for $T = A_T(\cdot) + c_T \in \mathcal{T}$ and $n \in \mathbb{Z}^d$,*

$$|h_{T,n}(x)| \leq C_N |T|^{1/2} (1 + \|A_T(x - na_T)\|_2)^{-N}, \quad \forall x \in \mathbb{R}^d.$$

Proof. We will use the fact that

$$u(\xi)^N = (1 + \|\xi\|_2)^N \asymp \sum_{|\beta|\leq N} \left|\xi^\beta\right|, \quad \xi \in \mathbb{R}^d, \tag{B.7}$$

for any $N \in \mathbb{N}$ with $\beta \in \mathbb{N}_0^d$. Let $\hat{g}_T(\xi) := \hat{h}_T(T\xi)$ such that $\operatorname{supp}(\hat{g}_T) \subset Q$ for all $T \in \mathcal{T}$. Using (B.7) we get

$$\begin{aligned}|g_T(x)| &\leq C_1(1 + \|x\|_2)^{-N} \sum_{|\beta|\leq N} \left|x^\beta g_T(x)\right| \\ &= C_1(1 + \|x\|_2)^{-N} \sum_{|\beta|\leq N} \left|\mathcal{F}^{-1}\left[\partial_\xi^\beta \hat{g}_T\right](x)\right| \\ &\leq C_1(1 + \|x\|_2)^{-N} \sum_{|\beta|\leq N} \int_{\mathbb{R}^d} \left|\partial_\xi^\beta \hat{g}_T(\xi)\right| d\xi, \quad x \in \mathbb{R}^d.\end{aligned}$$

Applying (B.6) we may continue and write

$$|g_T(x)| \leq C_2 a_T^{d/2} (1 + \|x\|_2)^{-N} \sum_{|\beta|\leq N} \int_{\mathbb{R}^d} \chi_Q(\xi) d\xi = C_3 a_T^{d/2} (1 + \|x\|_2)^{-N}. \tag{B.8}$$

Now, since $\varepsilon_T = C_*/a_T$ then

$$|T|\,|Q| = |Q_T| = \left(2\varepsilon_T + \frac{1}{a_T}\right)^d = (2C_* + 1)^d (a_T)^{-d}. \tag{B.9}$$

Hence, $a_T^{d/2} = C|T|^{-1/2}$ so (B.8) yields

$$|g_T(x)| \le C_4\,|T|^{-1/2}\,(1 + \|x\|_2)^{-N}, \quad x \in \mathbb{R}^d. \tag{B.10}$$

Using the fact that A_T is a diagonal matrix, we obtain the relationship

$$\begin{aligned} h_T(x) &= \int_{\mathbb{R}^d} \hat{h}_T(\xi) e^{2\pi i \xi \cdot x} d\xi = |T| \int_{\mathbb{R}^d} \hat{g}_T(u) e^{2\pi i (A_T u + c_T)\cdot x} du \\ &= e^{2\pi i c_T \cdot x}\,|T| \int_{\mathbb{R}^d} \hat{g}_T(u) e^{2\pi i u \cdot A_T x} du = e^{2\pi i c_T \cdot x}\,|T|\, g_T(A_T x), \quad x \in \mathbb{R}^d. \end{aligned} \tag{B.11}$$

Combining (B.11) and (B.10) we arrive at

$$\begin{aligned} |h_{T,n}(x)| &= |h_T(x - n a_T)| = |T|\,|g_T(A_T(x - n a_T))| \\ &\le C_4\,|T|^{1/2}\,(1 + \|A_T(x - n a_T)\|_2)^{-N}, \quad x \in \mathbb{R}^d. \end{aligned}$$

This proves the proposition. □

As a direct consequence of Proposition 4.1 we can prove the following lemma.

Lemma 4.2. *For $0 < p < \infty$, we have*

$$\sup_{x \in \mathbb{R}^d} \left\{ \left\| \{h_{T,n}(x)\}_{n \in \mathbb{Z}^d} \right\|_{\ell^p} \right\} \le C\,|T|^{1/2}, \quad \text{and} \tag{B.12}$$

$$\sup_{n \in \mathbb{Z}^d} \|h_{T,n}\|_{L^p} \le C'\,|T|^{1/2 - 1/p}, \tag{B.13}$$

with constants $C, C' > 0$ independent of $T \in \mathcal{T}$.

Proof. We will use the fact that

$$\int_{\mathbb{R}^d} u(\xi)^{-m} d\xi = \int_{\mathbb{R}^d} (1 + \|\xi\|_2)^{-m}\, d\xi < \infty, \tag{B.14}$$

for any $m > d$. Choosing $N > d/p$ in Proposition 4.1, then (B.14) yields

$$\begin{aligned} \left\| \{h_{T,n}(x)\}_{n \in \mathbb{Z}^d} \right\|_{\ell^p} &\le C_1\,|T|^{1/2} \left(\sum_{n \in \mathbb{Z}^d} (1 + \|A_T(x - n a_T)\|_2)^{-Np} \right)^{1/p} \\ &\le C_2\,|T|^{1/2}\,(a_T^d\,|T|)^{-1/p}. \end{aligned} \tag{B.15}$$

According to (B.9), $a_T^d = C|T|^{-1}$, which inserted into (B.15) yields (B.12). To show (B.13), we again let $N > d/p$ in Proposition 4.1, so (B.14) yields

$$\|h_{T,n}\|_{L^p} \leq C_1 |T|^{1/2} \left(\int_{\mathbb{R}^d} (1 + \|A_T(x - na_T)\|_2)^{-Np} dx \right)^{1/p} \leq C_2 |T|^{1/2-1/p}.$$

This proves (B.6). □

In the next section we use the painless NSGF $\{h_{T,n}\}_{T,n}$ to prove a complete characterization of the corresponding decomposition space $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$.

5 Characterization of decomposition spaces

The main result of this section is the characterization given in Theorem 5.1. To prove this result, we follow the approach taken in [3] where the authors proved a similar result for a certain type of tight frames for $\mathbb{R}^d$ (see [3, Proposition 3]). Since the frames we consider are not assumed to be tight we need to modify the arguments given in [3]. We start with the following observations.

Lemma 5.1. *For $0 < p < \infty$, the Fourier multiplier*

$$\psi_T^h(D)f := \mathcal{F}^{-1}\left(\psi_T^h \mathcal{F} f\right) := \mathcal{F}^{-1}\left(\frac{\psi_T}{\sum_{l\in\widetilde{T}} \frac{1}{a_l^d} \left|\hat{h}_l\right|^2} \mathcal{F} f \right) \tag{B.16}$$

is bounded on the band-limited functions in $L^p(\mathbb{R}^d)$ uniformly in $T \in \mathcal{T}$. Further,

$$\sup_{x\in\mathbb{R}^d} \left\{ \left\| \left\{ \psi_T^h(D) h_{T',n} \right\}_{n\in\mathbb{Z}^d} \right\|_{\ell^p} \right\} \leq C |T|^{1/2}, \quad T \in \mathcal{T}, \quad T' \in \widetilde{T}, \tag{B.17}$$

with a constant $C > 0$ independent of $T \in \mathcal{T}$.

Proof. Let $\psi_T^{h'}(\xi) := \psi_T^h(T(\xi))$. For $N > d/p$, (B.14) and (B.7) imply

$$\begin{aligned} \left\|\mathcal{F}^{-1}\psi_T^{h'}\right\|_{L^p} &\leq C_1 \left\| u(\cdot)^N \mathcal{F}^{-1}\psi_T^{h'} \right\|_{L^\infty} \leq C_2 \sum_{|\beta|\leq N} \left\| (\cdot)^\beta \mathcal{F}^{-1}\psi_T^{h'} \right\|_{L^\infty} \\ &= C_2 \sum_{|\beta|\leq N} \left\| \mathcal{F}^{-1}\left(\partial^\beta \psi_T^{h'}\right) \right\|_{L^\infty} \leq C_2 \sum_{|\beta|\leq N} \left\| \partial^\beta \psi_T^{h'} \right\|_{L^1}. \end{aligned} \tag{B.18}$$

Since $\varepsilon_T = C_*/a_T$, the chain rule yields

$$\partial^\beta \psi_T^{h'}(\xi) = \left(\partial^\beta \psi_T^h\right)(T\xi) \left(2\varepsilon_T + \frac{1}{a_T}\right)^{|\beta|} = C a_T^{-|\beta|} \left(\partial^\beta \psi_T^h\right)(T\xi). \tag{B.19}$$

For estimating $\partial^\beta \psi_T^h$ we use the quotient rule. Because all derivatives of ψ_T are bounded according to Proposition 2.1(4), we need only to consider the derivatives of the denominator of ψ_T^h. The sum in the denominator consists of at most n_0 terms and for each term in the sum, the chain rule and Definition 3.1(1)-(2) imply an upper bound of $Ca_T^{|\beta|}$. Therefore, (B.19) yields $|\partial^\beta \psi_T^{h'}(\xi)| \leq C'\chi_Q(\xi)$, since $\text{supp}(\psi_T^{h'}) \subset Q$ for all $T \in \mathcal{T}$. Combing this with (B.18) we get $\|\mathcal{F}^{-1}\psi_T^{h'}\|_{L^p} \leq C_3$. It now follows from Lemma A.2 page 81 that $f \to \mathcal{F}^{-1}(\psi_T^{h'}\mathcal{F}f)$ defines a bounded operator on the band-limited functions in $L^p(\mathbb{R}^d)$ uniformly in $T \in \mathcal{T}$. Finally, applying [3, Lemma 1] we obtain the same statement for $\psi_T^h(D)$.

We now prove (B.17). Repeating the arguments from the proof of Proposition 4.1 (using (B.3) and Definition 3.1(1)) we can prove the same decay property for $\psi_T^h(D)h_{T',n}$. The result therefore follows from the arguments in the proof of (B.12). $\square$

The statement in Theorem 5.1 follows directly once we have proven the following technical lemma. We use the notation $\widetilde{\psi}_T := \sum_{T'\in\widetilde{T}} \psi_{T'}$.

Lemma 5.2. *Given $f \in \mathcal{S}(\mathbb{R}^d)$ and $0 < p < \infty$. For all $T \in \mathcal{T}$,*

$$\left\|\{\langle f, h_{T,n}\rangle\}_{n\in\mathbb{Z}^d}\right\|_{\ell^p} \leq C\,|T|^{1/p-1/2}\left\|\widetilde{\psi}_T(D)f\right\|_{L^p}, \quad \textit{and} \tag{B.20}$$

$$\|\psi_T(D)f\|_{L^p} \leq C'\,|T|^{1/2-1/p}\sum_{T'\in\widetilde{T}}\left\|\{\langle f, h_{T',n}\rangle\}_{n\in\mathbb{Z}^d}\right\|_{\ell^p}, \tag{B.21}$$

with constants $C, C' > 0$ independent of $T \in \mathcal{T}$.

Proof. The proof of (B.20) follows directly from (B.12) and the arguments for the first part of the proof for [3, Lemma 2]. To prove (B.21) we first assume $p \leq 1$ and note

$$\begin{aligned}\|\psi_T(D)f\|_{L^p} &\leq C_1 \sum_{T'\in\widetilde{T}}\sum_{n\in\mathbb{Z}^d} |\langle f, h_{T',n}\rangle|\,\|\psi_T(D)\tilde{h}_{T',n}\|_{L^p} \\ &\leq C_2 \sum_{T'\in\widetilde{T}}\left(\sum_{n\in\mathbb{Z}^d}|\langle f, h_{T',n}\rangle|^p\,\|\psi_T(D)\tilde{h}_{T',n}\|_{L^p}^p\right)^{1/p},\end{aligned} \tag{B.22}$$

with $\{\tilde{h}_{T,n}\}_{T,n}$ being the dual frame given in (B.4) page 66. Applying (B.16) and (B.13) this proves (B.21) for the case $p \leq 1$. For $p > 1$, we note that

Hölders inequality (with p' being the conjugate index of p) yields

$$
\begin{aligned}
&\left\| \sum_{n\in\mathbb{Z}^d} \langle f, h_{T',n}\rangle\, \psi_T(D)\tilde{h}_{T',n} \right\|_{L^p}^p \\
&\leq \int_{\mathbb{R}^d} \sum_{n\in\mathbb{Z}^d} \left|\langle f, h_{T',n}\rangle\right|^p \left|\psi_T(D)\tilde{h}_{T',n}(x)\right| \left(\sum_{n'\in\mathbb{Z}^d} \left|\psi_T(D)\tilde{h}_{T',n'}(x)\right| \right)^{p/p'} dx \\
&\leq C_1 \left|T\right|^{p/2p'-1/2} \sum_{n\in\mathbb{Z}^d} \left|\langle f, h_{T',n}\rangle\right|^p ,
\end{aligned}
$$

according to Lemma 5.1 and (B.13). Taking the p'th root on both sides and applying (B.22) finishes the proof of (B.21) for $p > 1$. $\square$

Using the notation of [3] we define L^p-normalized atoms $h^p_{T,n} := |T|^{1/2-1/p}\, h_{T,n}$, for all $T \in \mathcal{T}$, $n \in \mathbb{Z}^d$ and $0 < p < \infty$. We also define the coefficient space $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ as the set of coefficients $\{c_{T,n}\}_{T\in\mathcal{T},n\in\mathbb{Z}^d} \subset \mathbb{C}$ satisfying

$$
\left\| \{c_{T,n}\}_{T\in\mathcal{T},n\in\mathbb{Z}^d} \right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})} := \left\| \left\{ \left\| \{c_{T,n}\}_{n\in\mathbb{Z}^d} \right\|_{\ell^p} \right\}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}} < \infty.
$$

Combining Lemma 5.2 with the fact that $\mathcal{S}(\mathbb{R}^d)$ is dense in $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ we obtain a characterization similar to that of [3, Proposition 3].

Theorem 5.1. *For $s \in \mathbb{R}$ and $0 < p, q < \infty$ we have the equivalence*

$$
\|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \asymp \left\| \left\{ \left\langle f, h^p_{T,n} \right\rangle \right\}_{T\in\mathcal{T},n\in\mathbb{Z}^d} \right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})},
$$

for all $f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$.

Remark 5.1. The characterization in Theorem 5.1 differs from the one given in [3, Proposition 3] in two ways. In [3] the frame elements are obtained directly from the structured covering such that the resulting system forms a tight frame. In our framework we take the "reverse" approach and explicitly state sufficient conditions which guarantee the existence of a compatible decomposition space for a given NSGF (cf. Definition 3.1). More importantly, we show that the assumption on tightness of the frame can be replaced with the structured expression for the dual frame given in (B.4) page 66.

In the next section we use the characterization given in Theorem 5.1 to prove that $\{h^p_{T,n}\}_{T,n}$ forms a Banach frame for $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ with respect to $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ for $s \in \mathbb{R}$ and $0 < p, q < \infty$.

6 Banach frames for decomposition spaces

Let us start by giving the general definition of a Banach frame [26, 27]. Traditionally, Banach frames are only defined for Banach spaces but we will also use the concept for (quasi-)Banach spaces.

Definition 6.1 (Banach Frame)**.** Let X be a (quasi-)Banach space and let X_d be an associated (quasi-)Banach sequence space on $\mathbb{N}$. A Banach frame for X, with respect to X_d, is a sequence $\{y_n\}_{n\in\mathbb{N}}$ in the dual space X', such that

1. The coefficient operator $C_X : f \to \{\langle f, y_n\rangle\}_{n\in\mathbb{N}}$ is bounded from X into X_d.

2. Norm equivalence:
$$\|f\|_X \asymp \|\{\langle f, y_n\rangle\}_{n\in\mathbb{N}}\|_{X_d}, \quad \forall f \in X.$$

3. There exists a bounded operator R_{X_d} from X_d onto X, called a reconstruction operator, such that
$$R_{X_d} C_X f = R_{X_d}\left(\{\langle f, y_n\rangle\}_{n\in\mathbb{N}}\right) = f, \quad \forall f \in X.$$

Remark 6.1. We will actually prove that $\{h^p_{T,n}\}_{T,n}$ forms an atomic decomposition [5, 18, 19] for $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ as the reconstruction operator takes the form $f = \sum_{T,n}\langle f, h^p_{T,n}\rangle x_{T,n}$ with $\{x_{T,n}\} \subset D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ (see Theorem 6.1 below).

In order to show that that $\{h^p_{T,n}\}_{T,n}$ forms a Banach frame for $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$, we first note that
$$\{h^p_{T,n}\}_{T\in\mathcal{T}, n\in\mathbb{Z}^d} \subset \mathcal{S}(\mathbb{R}^d) \subset D'(\mathcal{Q}, L^p, \ell^q_{\omega^s})$$
as required by Definition 6.1. Furthermore, the equivalence in Theorem 5.1 implies that Definition 6.1(2) is satisfied and the corresponding proof reveals that Definition 6.1(1) is satisfied. What remains to be shown is the existence of a bounded reconstruction operator such that Definition 6.1(3) holds. For $\{c_{T,n}\}_{T,n} \in d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$, we define the reconstruction operator as
$$R_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})}\left(\{c_{T,n}\}_{T,n}\right) = \sum_{T\in\mathcal{T}, n\in\mathbb{Z}^d} c_{T,n}\,|T|^{1/p-1/2}\,\tilde{h}_{T,n}, \tag{B.23}$$
with $\{\tilde{h}_{T,n}\}_{T\in\mathcal{T}, n\in\mathbb{Z}^d}$ being the dual frame given in (B.4) page 66. We now provide the main result of this section.

Theorem 6.1. *Given $s \in \mathbb{R}$ and $0 < p, q < \infty$, $\{h^p_{T,n}\}_{T\in\mathcal{T},n\in\mathbb{Z}^d}$ forms a Banach frame for $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$. Furthermore, we have the expansions*

$$f = \sum_{T\in\mathcal{T},n\in\mathbb{Z}^d} \langle f, h_{T,n}\rangle \tilde{h}_{T,n}, \quad \forall f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s}), \tag{B.24}$$

with unconditional convergence.

Proof. Let R and C denote the reconstruction- and coefficient operator, respectively. We first prove that R is bounded from $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ onto $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$. For $\{c_{T,n}\}_{T,n} \in d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ we let $g := R(\{c_{T,n}\}_{T,n})$. For $T \in \mathcal{T}$, Lemma 5.1 implies

$$\begin{aligned}\|\psi_T(D)g\|_{L^p} &= \left\| \mathcal{F}^{-1}\left(\frac{\psi_T \cdot \widetilde{\psi}_T}{\sum_{l\in\widetilde{T}} \frac{1}{a_l^d}\left|\hat{h}_l\right|^2} \cdot \sum_{T'\in\mathcal{T},n\in\mathbb{Z}^d} c_{T',n}\left|T'\right|^{1/p-1/2}\hat{h}_{T',n}\right)\right\|_{L^p} \\ &\le C_1 \left\| \widetilde{\psi}_T(D)\left(\sum_{T'\in\mathcal{T},n\in\mathbb{Z}^d} c_{T',n}\left|T'\right|^{1/p-1/2} h_{T',n}\right)\right\|_{L^p}. \end{aligned}\tag{B.25}$$

Repeating the arguments from the proof of [3, Lemma 4] we can show that

$$\left\|\sum_{T\in\mathcal{T},n\in\mathbb{Z}^d} c_{T,n}\left|T\right|^{1/p-1/2} h_{T,n}\right\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \le C\left\|\{c_{T,n}\}_{T,n}\right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})}. \tag{B.26}$$

Applying (B.2) page 63 to (B.25) and then using (B.26) we get

$$\begin{aligned}\|g\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} &\le C_2 \left\|\sum_{T\in\mathcal{T},n\in\mathbb{Z}^d} c_{T,n}\left|T\right|^{1/p-1/2} h_{T,n}\right\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \\ &\le C_3 \left\|\{c_{T,n}\}_{T\in\mathcal{T},n\in\mathbb{Z}^d}\right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})}. \end{aligned}\tag{B.27}$$

This proves that R is bounded from $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ onto $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$. Let us now show the unconditional convergence of (B.24). Given $f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$, we can find a sequence $\{f_k\}_{k\ge 1}$, with $f_k \in \mathcal{S}(\mathbb{R}^d)$ for all $k \ge 1$, such that $f_k \to f$ in $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ as $k \to \infty$. Furthermore, since $\{h_{T,n}\}_{T,n}$ forms a frame for $L^2(\mathbb{R}^d)$, for each $k \ge 1$ we have the expansion

$$f_k = \sum_{T\in\mathcal{T},n\in\mathbb{Z}^d} \langle f_k, h_{T,n}\rangle \tilde{h}_{T,n} = RC(f_k),$$

with unconditional convergence. Since $RC : D(\mathcal{Q}, L^p, \ell^q_{\omega^s}) \to D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ is continuous, letting $k \to \infty$ yields

$$f = RC(f) = \sum_{T\in\mathcal{T},n\in\mathbb{Z}^d} \langle f, h_{T,n}\rangle \tilde{h}_{T,n}. \tag{B.28}$$

Given $\varepsilon > 0$, (B.27) implies that we can find a *finite* subset $F_0 \subset \mathcal{T} \times \mathbb{Z}^d$, such that

$$\left\| f - \sum_{(T,n)\in F} \langle f, h_{T,n}\rangle \, \tilde{h}_{T,n} \right\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \leq C \left\| \{\langle f, h_{T,n}\rangle\}_{(T,n)\notin F} \right\|_{d(\mathcal{Q},L^p,\ell^q_{\omega^s})} < \varepsilon,$$

for all finite sets $F \supseteq F_0$. According to [27, Proposition 5.3.1], this property is equivalent to unconditional convergence. □

We close this section by discussing the implications of the achieved results. According to Theorem 5.1 and Theorem 6.1, every $f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ has an expansion of the form

$$f = \sum_{T\in\mathcal{T}, n\in\mathbb{Z}^d} \left\langle f, h^p_{T,n} \right\rangle |T|^{1/p-1/2} \, \tilde{h}_{T,n}, \quad \text{with}$$

$$\|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \asymp \left\| \left\{ \left\langle f, h^p_{T,n} \right\rangle \right\}_{T,n} \right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})}.$$

Now, assume there exists another set of reconstruction coefficients $\{c_{T,n}\}_{T,n} \in d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ which is sparser than $\{\langle f, h^p_{T,n}\rangle\}_{T,n}$ when sparseness is measured by the $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$-norm. Since the reconstruction operator R is bounded we get

$$\begin{aligned}
\|\{c_{T,n}\}_{T,n}\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})} &\leq \left\| \left\{ \left\langle f, h^p_{T,n} \right\rangle \right\}_{T,n} \right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})} \leq C_1 \|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \\
&= C_1 \|R(\{c_{T,n}\}_{T,n})\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} \\
&\leq C_2 \|\{c_{T,n}\}_{T,n}\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})}.
\end{aligned}$$

We conclude that the canonical coefficients $\{\langle f, h^p_{T,n}\rangle\}_{T,n}$ are (up to a constant) the sparsest possible choice for expanding f as

$$f = \sum_{T\in\mathcal{T}, n\in\mathbb{Z}^d} c_{T,n} \, |T|^{1/p-1/2} \, \tilde{h}_{T,n},$$

when sparseness of the coefficients is measured by the $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$-norm. Furthermore, $f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ if and only if f permits a sparse expansion relative to the dictionary $\{|T|^{1/p-1/2}\widetilde{h}_{T,n}\}_{T,n}$.

7 Application to nonlinear approximation theory

In this section we provide the link to nonlinear approximation theory. An important property of the sparse expansions obtained in Theorem 6.1 is that we can obtain a good compression by simply thresholding the coefficients from the expansion. As mentioned in the introduction, NSGFs can create adaptive time-frequency representations as opposed to standard Gabor frames. Such adaptive representations can be constructed to fit the particular nature of a given signal, thereby producing a more precise (and hopefully sparser) time-frequency representation. In particular, NSGFs have proven to be useful in connection with music signals. For instance, in [1, 11] the authors use NSGFs to construct an invertible constant-Q transform with good frequency resolution at the lower frequencies and good time resolution at the higher frequencies. Such a time-frequency resolution is often more natural for music signals than the uniform resolution provided by Gabor frames.

The main result of this section is given in (B.30) below. The corresponding proof follows directly from the results obtained in Sections 5 and 6 together with standard arguments from nonlinear approximation theory [23]. Let $f \in D(\mathcal{Q}, L^\tau, \ell^\tau_{\omega^s})$, with $0 < \tau < \infty$, and let $0 < p < \infty$ satisfy $\alpha := 1/\tau - 1/p > 0$. Write the frame expansion of f with respect to the L^p-normalized coefficients

$$f = \sum_{T \in \mathcal{T}, n \in \mathbb{Z}^d} \left\langle f, h^p_{T,n} \right\rangle |T|^{1/p-1/2}\, \tilde{h}_{T,n}. \tag{B.29}$$

Let $\{\theta_m\}_{m \in \mathbb{N}}$ be a decreasing rearrangement of the frame coefficients and let f_N be the N-term approximation to f obtained by extracting the coefficients in (B.29) corresponding to the N largest coefficients $\{\theta_m\}_{m=1}^N$. Then, we can prove the existence of $C > 0$ such that for $f \in D(\mathcal{Q}, L^\tau, \ell^\tau_{\omega^s})$ and $N \in \mathbb{N}$,

$$\|f - f_N\|_{D(\mathcal{Q}, L^p, \ell^p_{\omega^s})} \leq C N^{-\alpha} \|f\|_{D(\mathcal{Q}, L^\tau, \ell^\tau_{\omega^s})}. \tag{B.30}$$

In other words, for $f \in D(\mathcal{Q}, L^\tau, \ell^\tau_{\omega^s})$ we obtain good approximations in $D(\mathcal{Q}, L^p, \ell^p_{\omega^s})$ by thresholding the L^p-normalized frame coefficients. We note that for $0 < \tau < 2$ we obtain good approximations in $L^2(\mathbb{R}^d)$ with respect to the original coefficients $\{\langle f, h_{T,n} \rangle\}_{T,n}$.

We now explain the obtained results in the general framework of Jackson- and Bernstein inequalities [8]. Let $\mathcal{D}$ denote the dictionary $\{|T|^{1/p-1/2} \tilde{h}_{T,n}\}_{T,n}$ and define the nonlinear set of all linear combinations of at most N elements from $\mathcal{D}$ as

$$\Sigma_N(\mathcal{D}) := \left\{ \sum_{T,n \in \Delta} c_{T,n} |T|^{1/p-1/2}\, \tilde{h}_{T,n} \;\middle|\; \#\Delta \leq N \right\}.$$

For any $f \in D(\mathcal{Q}, L^p, \ell^p_{\omega^s})$, the error of best N-term approximation to f is

$$\sigma_N(f, \mathcal{D}) := \inf_{h \in \Sigma_N(\mathcal{D})} \|f - h\|_{D(\mathcal{Q}, L^p, \ell^p_{\omega^s})}.$$

Since $f_N \in \Sigma_N(\mathcal{D})$, (B.30) yields

$$\sigma_N(f, \mathcal{D}) \leq C N^{-\alpha} \|f\|_{D(\mathcal{Q}, L^\tau, \ell^\tau_{\omega^s})}.$$

This is a so-called Jackson inequality for nonlinear N-term approximation with $\mathcal{D}$. It provides us with an upper bound for the error obtained by approximating f with the best possible choice of linear combinations of at most N elements from the dictionary. The converse inequality is called a Bernstein inequality and is in general much more difficult to obtain for redundant systems [24]. The existence of a Bernstein inequality would provide us with a lower bound and hence a full characterization of the error of best N-term approximation to f with respect to the dictionary $\mathcal{D}$. However, for this particular system (and for many other redundant systems), the existence of a Bernstein inequality is still an open question.

Acknowledgements

We thank the two anonymous reviewers for their constructive comments on the original manuscript. Their valuable suggestions have helped improve the manuscript considerably.

A Proof of Theorem 2.1

Proof. To simplify notation we let $D^s_{p,q} := D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$. Let us first prove Theorem 2.1(1). Allowing the extension $q = \infty$, and repeating the arguments from the proof of [4, Proposition 5.7], we can show that

$$D^{s+\varepsilon}_{p,\infty} \hookrightarrow D^s_{p,q} \hookrightarrow D^s_{p,\infty}, \quad \varepsilon > d/q,$$

for any $s \in \mathbb{R}$ and $0 < p < \infty$ using Definition 2.2(4). It therefore suffice to show that $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{p,\infty} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ for any $s \in \mathbb{R}$ and $0 < p < \infty$. For $N \in \mathbb{N}$, we define semi-norms on $\mathcal{S}(\mathbb{R}^d)$ by

$$p_N(g) := \sup_{\xi \in \mathbb{R}^d} \left\{ u(\xi)^N \sum_{|\beta| \leq N} \left| \partial^\beta \hat{g}(\xi) \right| \right\}, \quad g \in \mathcal{S}(\mathbb{R}^d),$$

with $u(\xi) = 1 + \|\xi\|_2$ as usual. Following the approach in [4, Page 149], and applying Proposition 2.1(4), we get

$$\|f\|_{D^s_{p,\infty}} \leq C p_N(f) \quad \text{and} \quad \|f\|_{D^s_{1,1}} \leq C' p_{N'}(f),$$

for sufficiently large N and N'. This proves that $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{p,\infty}$ and $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{1,1}$. To show that $D^s_{p,\infty} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ we need to take a different approach than in [4]. Setting $\widetilde{\psi}_T := \sum_{T' \in \widetilde{T}} \psi_{T'}$, we first note that for $f \in D^s_{p,\infty}$ and $\varphi \in \mathcal{S}(\mathbb{R}^d)$,

$$|\langle f,\varphi\rangle| \le \sum_{T\in\mathcal{T}} \left\|\psi_T(D)f\widetilde{\psi}_T(D)\varphi\right\|_{L^1} \le \sum_{T\in\mathcal{T}} \|\psi_T(D)f\|_{L^\infty} \left\|\widetilde{\psi}_T(D)\varphi\right\|_{L^1}.$$

Using Lemma A.1 below (with $g = \mathcal{F}^{-1}\{\psi_T\hat{f}(T\xi)\}$) we thus get

$$\begin{aligned}
|\langle f,\varphi\rangle| &\le C_1 \sum_{T\in\mathcal{T}} |T|^{1/p} \|\psi_T(D)f\|_{L^p} \left\|\widetilde{\psi}_T(D)\varphi\right\|_{L^1} \\
&\le C_1 \|f\|_{D^s_{p,\infty}} \sum_{T\in\mathcal{T}} |T|^{1/p}\, \omega_T^{-s} \left\|\widetilde{\psi}_T(D)\varphi\right\|_{L^1} \\
&\le C_2 \|f\|_{D^s_{p,\infty}} \left\| \left\{ \left\|\widetilde{\psi}_T(D)\varphi\right\|_{L^1} \right\}_{T\in\mathcal{T}} \right\|_{\ell^1_{\omega^{\gamma/p-s}}}, \qquad \text{(B.31)}
\end{aligned}$$

since $|T| = |Q|^{-1}|Q_T| \le |Q|^{-1}\omega_T^{\gamma}$ according to Definition 2.2(5). Applying (B.2) page 63 we may continue on (B.31) and write

$$|\langle f,\varphi\rangle| \le C_3 \|f\|_{D^s_{p,\infty}} \|\varphi\|_{D^{\gamma/p-s}_{1,1}} \le C_4 \|f\|_{D^s_{p,\infty}}\, p_N(\varphi),$$

for sufficiently large N since $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{1,1}$. We conclude that $D^s_{p,\infty} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ which proves Theorem 2.1(1).

The proof of Theorem 2.1(2) follows directly from Theorem 2.1(1) and the arguments in [4, Page 150].

To prove Theorem 2.1(3) we let $f \in D^s_{p,q}$ and choose $I \in C_c^\infty(\mathbb{R}^d)$ with $0 \le I(\xi) \le 1$ and $I(\xi) \equiv 1$ in a neighborhood of $\xi = 0$. Also, we define $(\widetilde{f})\widehat{\ } := I\hat{f}$ and

$$\widetilde{f}_\varepsilon := \mathcal{F}^{-1}\left\{ \varphi_\varepsilon * \left(\widetilde{f}\right)\widehat{\ } \right\} \in \mathcal{S}(\mathbb{R}^d),$$

with $\varphi_\varepsilon(\xi) := \varepsilon^{-d}\varphi(\xi/\varepsilon)$ and φ being a compactly supported mollifier. Since $\mathrm{supp}(I)$ is compact, we may choose a *finite* subset $T^* \subset \mathcal{T}$ with $\mathrm{supp}(I) \subset \cup_{T\in T^*} Q_T$ and $\sum_{T\in T^*} \psi_T(\xi) \equiv 1$ on $\mathrm{supp}(I)$. Using Lemma A.2 below we obtain

$$\begin{aligned}
\|\widetilde{f}\|_{L^p} &= \left\| \mathcal{F}^{-1} I \mathcal{F}\left(\mathcal{F}^{-1}\left(\sum_{T\in T^*} \psi_T \cdot \hat{f} \right)\right) \right\|_{L^p} \\
&\le C \sum_{T\in T^*} \left\| \mathcal{F}^{-1} I \right\|_{L^{\tilde{p}}} \|\psi_T(D)f\|_{L^p} < \infty,
\end{aligned}$$

with $\tilde{p} = \min\{1,p\}$. The dominated convergence theorem thus yields

$$\left\| \widetilde{f} - \widetilde{f}_\varepsilon \right\|_{D^s_{p,q}} \le C \left\| \left\{ \left\| \widetilde{f} - \widetilde{f}_\varepsilon \right\|_{L^p} \right\}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}} \to 0, \quad \text{as } \varepsilon \to 0,$$

so the proof is done if we can show that $\|f - \widetilde{f}\|_{D^s_{p,q}}$ can be made arbitrary small by choosing $\widetilde{f}$ appropriately. To show this, we define the set

$T_\circ := \{T \in \mathcal{T} \mid I(\xi) \equiv 1 \text{ on } \operatorname{supp}(\psi_T)\}$. Denoting the complement $T_\circ^c$, Lemma A.2 below yields

$$\begin{aligned} \left\| f - \widetilde{f} \right\|_{D^s_{p,q}}^q &= \sum_{T \in T_\circ^c} \omega_T^{sq} \left\| \mathcal{F}^{-1} \left(\psi_T \left(\hat{f} - I\hat{f} \right) \right) \right\|_{L^p}^q \\ &\leq C_1 \sum_{T \in T_\circ^c} \omega_T^{sq} \left(\|\psi_T(D) f\|_{L^p} + \left\| \mathcal{F}^{-1} I \mathcal{F} \left(\psi_T(D) f \right) \right\|_{L^p} \right)^q \\ &\leq C_2 \sum_{T \in T_\circ^c} \omega_T^{sq} \|\psi_T(D) f\|_{L^p}^q . \end{aligned}$$

Finally, since $f \in D^s_{p,q}$ we can choose $\operatorname{supp}(I)$ large enough, such that $\|f - \widetilde{f}\|_{D^s_{p,q}} < \varepsilon$, for any given $\varepsilon > 0$. This proves Theorem 2.1(3). □

In the proof of Theorem 2.1 we used the following two lemmas. A proof of Lemma A.1 can be found in [3, Lemma 3] and a proof of Lemma A.2 can be found in [41, Proposition 1.5.1].

Lemma A.1. *Let $g \in L^p(\mathbb{R}^d)$ and $\operatorname{supp}(\hat{g}) \subset \Gamma$, with $\Gamma \subset \mathbb{R}^d$ compact. Given an invertible affine transformation T, let $\hat{g}_T(\xi) := \hat{g}(T^{-1}\xi)$. Then for $0 < p \leq q \leq \infty$,*

$$\|g_T\|_{L_q} \leq C \, |T|^{1/p - 1/q} \, \|g_T\|_{L_p} ,$$

for a constant C independent of T.

Lemma A.2. *Let Ω and Γ be compact subsets of $\mathbb{R}^d$. Let $0 < p \leq \infty$ and $\tilde{p} = \min\{1, p\}$. Then there exists a constant C such that*

$$\left\| \mathcal{F}^{-1} M \mathcal{F} f \right\|_{L^p} \leq C \left\| \mathcal{F}^{-1} M \right\|_{L^{\tilde{p}}} \|f\|_{L^p}$$

for all $f \in L^p(\mathbb{R}^d)$ with $\operatorname{supp}(\hat{f}) \subset \Omega$ and all $\mathcal{F}^{-1} M \in L^{\tilde{p}}(\mathbb{R}^d)$ with $\operatorname{supp}(M) \subset \Gamma$.

References

[1] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco. Theory, implementation and applications of nonstationary Gabor frames. *J. Comput. Appl. Math.*, 236(6):1481–1496, 2011.

[2] L. Borup and M. Nielsen. Banach frames for multivariate α-modulation spaces. *J. Math. Anal. Appl.*, 321(2):880–895, 2006.

[3] L. Borup and M. Nielsen. Frame decomposition of decomposition spaces. *J. Fourier Anal. Appl.*, 13(1):39–70, 2007.

[4] L. Borup and M. Nielsen. On anisotropic Triebel-Lizorkin type spaces, with applications to the study of pseudo-differential operators. *J. Funct. Spaces Appl.*, 6(2):107–154, 2008.

[5] P. Casazza, D. Han, and D. Larson. Frames for Banach spaces. *Contemporary Mathematics*, 247:149–182, 1999.

[6] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], second edition, 2016.

[7] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5):1271–1283, 1986.

[8] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[9] M. Dörfler. Time-frequency analysis for music signals: A mathematical approach. *Journal of New Music Research*, 30(1):3–12, 2001.

[10] M. Dörfler. Quilted Gabor frames—a new concept for adaptive time-frequency representation. *Adv. in Appl. Math.*, 47(4):668–687, 2011.

[11] M. Dörfler, N. Holighaus, T. Grill, and G. Velasco. Constructing an invertible constant-Q transform with nonstationary Gabor frames. In *n Proceedings of the 14th International Conference on Digital Audio Effects (DAFx 11)*, 2011.

[12] M. Dörfler and E. Matusiak. Nonstationary Gabor frames—existence and construction. *Int. J. Wavelets Multiresolut. Inf. Process.*, 12(3):1450032, 18, 2014.

[13] M. Dörfler and E. Matusiak. Nonstationary Gabor frames—approximately dual frames and reconstruction errors. *Adv. Comput. Math.*, 41(2):293–316, 2015.

[14] H. G. Feichtinger. Banach convolution algebras of Wiener type. In *Functions, series, operators, Vol. I, II (Budapest, 1980)*, volume 35 of *Colloq. Math. Soc. János Bolyai*, pages 509–524. North-Holland, Amsterdam, 1983.

[15] H. G. Feichtinger. Modulation spaces on locally compact abelian groups. Technical report, University of Vienna, 1983.

[16] H. G. Feichtinger. Banach spaces of distributions defined by decomposition methods. II. *Math. Nachr.*, 132:207–237, 1987.

[17] H. G. Feichtinger and P. Gröbner. Banach spaces of distributions defined by decomposition methods. I. *Math. Nachr.*, 123:97–120, 1985.

[18] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. I. *J. Funct. Anal.*, 86(2):307–340, 1989.

[19] H. G. Feichtinger and K. H. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions. II. *Monatsh. Math.*, 108(2-3):129–148, 1989.

[20] M. Frazier and B. Jawerth. Decomposition of Besov spaces. *Indiana Univ. Math. J.*, 34(4):777–799, 1985.

[21] H. Führ and F. Voigtlaender. Wavelet coorbit spaces viewed as decomposition spaces. *J. Funct. Anal.*, 269(1):80–154, 2015.

[22] R. Gribonval and M. Nielsen. Some remarks on non-linear approximation with Schauder bases. *East J. Approx.*, 7(3):267–285, 2001.

[23] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51–71, 2004.

[24] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. II. Inverse estimates. *Constr. Approx.*, 24(2):157–173, 2006.

[25] P. Gröbner. *Banachräume glatter Funktionen und Zerlegungsmethoden*. ProQuest LLC, Ann Arbor, MI, 1992. Thesis (Dr.natw.)–Technische Universität Wien (Austria).

[26] K. Gröchenig. Describing functions: atomic decompositions versus frames. *Monatsh. Math.*, 112(1):1–42, 1991.

[27] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[28] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–331, 2000.

[29] C. Heil. An introduction to weighted Wiener amalgams. *Wavelets and their Applications (Chennai, January 2002)*, pages 183–216, 2003.

[30] E. Hernández, D. Labate, and G. Weiss. A unified characterization of reproducing systems generated by a finite family. II. *J. Geom. Anal.*, 12(4):615–662, 2002.

[31] N. Holighaus. Structure of nonstationary Gabor frames and their dual systems. *Appl. Comput. Harmon. Anal.*, 37(3):442–463, 2014.

[32] N. Holighaus, M. Dörfler, G. A. Velasco, and T. Grill. A framework for invertible, real-time constant-Q transforms. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(4):775–785, April 2013.

[33] A. Holzapfel, G. A. Velasco, N. Holighaus, M. Dörfler, and A. Flexer. Advantages of nonstationary Gabor transforms in beat tracking. In *Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, MIRUM '11, pages 45–50, New York, NY, USA, 2011. ACM.

[34] F. Jaillet and B. Torrésani. Time-frequency jigsaw puzzle: adaptive multiwindow and multilayered Gabor expansions. *Int. J. Wavelets Multiresolut. Inf. Process.*, 5(2):293–315, 2007.

[35] M. S. Jakobsen and J. Lemvig. Reproducing formulas for generalized translation invariant systems on locally compact abelian groups. *Trans. Amer. Math. Soc.*, 368(12):8447–8480, 2016.

[36] G. Kutyniok and D. Labate. The theory of reproducing systems on locally compact abelian groups. *Colloq. Math.*, 106(2):197–220, 2006.

[37] M. Nielsen and K. N. Rasmussen. Compactly supported frames for decomposition spaces. *J. Fourier Anal. Appl.*, 18(1):87–117, 2012.

[38] W. J. Pielemeier, G. H. Wakefield, and M. H. Simoni. Time-frequency analysis of musical signals. *Proceedings of the IEEE*, 84(9):1216–1230, Sep 1996.

[39] H. Rauhut. Wiener amalgam spaces with respect to quasi-Banach spaces. *Colloq. Math.*, 109(2):345–362, 2007.

[40] A. Ron and Z. Shen. Generalized shift-invariant systems. *Constr. Approx.*, 22(1):1–45, 2005.

[41] H. Triebel. *Theory of function spaces*, volume 38 of *Mathematik und ihre Anwendungen in Physik und Technik [Mathematics and its applications in physics and technology]*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1983.

[42] P. J. Wolfe, S. J. Godsill, and M. Dörfler. Multi-Gabor dictionaries for audio time-frequency analysis. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 43–46, 2001.

[43] M. Zibulski and Y. Y. Zeevi. Analysis of multiwindow Gabor-type schemes by frame methods. *Appl. Comput. Harmon. Anal.*, 4(2):188–221, 1997.

Paper C

Nonlinear Approximation with Nonstationary Gabor Frames

Emil Solsbæk Ottosen and Morten Nielsen

The paper has been accepted for publication in
Advances in Computational Mathematics, 2017.

Abstract

We consider sparseness properties of adaptive time-frequency representations obtained using nonstationary Gabor frames (NSGFs). NSGFs generalize classical Gabor frames by allowing for adaptivity in either time or frequency. It is known that the concept of painless nonorthogonal expansions generalizes to the nonstationary case, providing perfect reconstruction and an FFT based implementation for compactly supported window functions sampled at a certain density. It is also known that for some signal classes, NSGFs with flexible time resolution tend to provide sparser expansions than can be obtained with classical Gabor frames. In this article we show, for the continuous case, that sparseness of a nonstationary Gabor expansion is equivalent to smoothness in an associated decomposition space. In this way we characterize signals with sparse expansions relative to NSGFs with flexible time resolution. Based on this characterization we prove an upper bound on the approximation error occurring when thresholding the coefficients of the corresponding frame expansions. We complement the theoretical results with numerical experiments, estimating the rate of approximation obtained from thresholding the coefficients of both stationary and nonstationary Gabor expansions.

1 Introduction

The field of Gabor theory [6, 19, 41] is concerned with representing signals as atomic decompositions using time-frequency localized atoms. The atoms are constructed as time-frequency shifts of a fixed window function, according to some lattice parameters, such that the resulting system constitutes a frame and, therefore, guarantees stable expansions [5, 31, 35]. Such frames are known under the name of Weyl-Heisenberg frames or Gabor frames and have been proven useful in a variety of applications [10, 22, 33]. The structure of Gabor frames implies a time-frequency resolution which depends only on the lattice parameters and the window function. In particular, the resolution is independent of the signal under consideration, which makes the corresponding implementation fast and easy to handle. The usage of a predetermined time-frequency resolution naturally raises the question of whether an improvement can be obtained by taking the signal class into consideration? This question has lead to many interesting approaches for constructing adaptive time-frequency representations [11, 27, 40, 42]. Unfortunately, for representations with resolution varying in *both* time and frequency there seems to be a trade-off between perfect reconstruction and fast implementation [30]. In this article, we therefore consider time-frequency representations with resolution varying in *either* time or frequency. The idea is to generalise the theory of painless nonorthogonal expansions [6] to the situation where multiple window functions are used along either the time- or the frequency axis. The resulting systems, which allow for perfect reconstruction and an FFT based implementation, are called painless generalised shift-invariant systems [25, 36] or painless nonstationary Gabor frames (painless NSGFs) [1, 26]. As already noted in [1], painless NSGFs tend to produce sparser representations than classical Gabor frames for certain classes of music signals. Sparseness of a time-frequency representation is desirable for several reasons, mainly because it may reduce the computational cost for manipulating and storing the coefficients [8, 18]. Additionally, many signal classes are characterized by some kind of sparseness in time or frequency and the corresponding signals are, therefore, best described by a sparse time-frequency representation. For such signals, the task of feature identification also benefits from a sparse representation as the particular characteristics of the signal becomes easier to identify.

In this article we consider sparseness properties of painless NSGFs with resolution varying in time. Whereas modulation spaces [14, 16, 22] have turned out to be the proper function spaces for analyzing sparseness properties of classical Gabor frames [23], we need a more general framework for the nonstationary case. A painless NSGF with flexible time resolution corresponds to a sampling grid which is irregular over time but regular over frequency for each fixed time point. We therefore search for a smoothness space which is compatible with a (more or less) arbitrary partition of the time domain. Such a flexibility can be provided by decomposition spaces, as introduced by Feichtinger and Gröbner in [15, 17]. Decomposition spaces may be viewed as a generalization of the classical Wiener amalgam spaces [13, 24] but with no assumption of an upper bound on the measure of the members of the partition. Another way of stating this is that decomposition spaces are constructed using bounded *admissible* partitions of unity [17] instead of bounded *uniform* partitions of unity [13]. The partitions we consider are obtained by applying a set of invertible affine transformations $\{A_k(\cdot) + c_k\}_{k\in\mathbb{N}}$ on a fixed set $Q \subset \mathbb{R}^d$ [2].

We use decomposition spaces to characterize signals with sparse expansions relative to painless NSGFs with flexible time resolution. We measure sparseness of an expansion by a mixed norm on the coefficients and show that the sparseness property implies an upper bound on the approximation error obtained by thresholding the expansion. Using the terminology from nonlinear approximation, such an upper bound is also known as a Jackson inequality [4, 8]. A similar characterization for classical Gabor frames using modulation spaces was proven by Gröchenig and Samarah in [23]. For the nonstationary case, we provided a characterization in [32] for painless NSGFs with flexible frequency resolution using decomposition spaces. A different approach to this problem is considered by Voigtlaender in [39], where the painless assumption is replaced with a more general analysis of the sampling parameter. The decomposition spaces considered in both [32] and [39] are based on partitions of the frequency domain, which is not a natural choice for NSGFs with flexible time resolution. In this article we consider decompositions of the time domain, which allow for compactly supported window functions sampled at a low density (compared to the general theory formulated in [39]). It is worth noting that there is a significant mathematical difference between decomposition spaces in time and in frequency.

The structure of this this article is as follows. In Section 2 we formally introduce decomposition spaces in time and prove several important properties of these spaces. Then, based on the ideas in [32], we show in Section 3 how to construct a suitable decomposition space for a given painless NSGF with flexible time resolution. In Section 4 we prove that the suitable decomposition space characterizes signals with sparse frame expansions and we provide an upper bound on the approximation rate occurring when thresholding the frame coefficients. Finally, in Section 5 we present the numerical results and in Section 6 we give the conclusions.

Let us now briefly go through our notation. By $\hat{f}(\xi) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i x\cdot\xi}dx$ we denote the Fourier transform with the usual extension to $L^2(\mathbb{R}^d)$. With $F \asymp G$ we mean that there exist two constants $0 < C_1, C_2 < \infty$ such that $C_1 F \leq G \leq C_2 F$. For two normed vector spaces X and Y, $X \hookrightarrow Y$ means that $X \subset Y$ and $\|f\|_Y \leq C\|f\|_X$ for some constant C and all $f \in X$. We say that a non-empty open set $\Omega' \subset \mathbb{R}^d$ is *compactly contained* in an open set $\Omega \subset \mathbb{R}^d$ if $\overline{\Omega'} \subset \Omega$ and $\overline{\Omega'}$ is compact. We call $\{x_i\}_{i\in\mathcal{I}} \subset \mathbb{R}^d$ a $\delta-$separated set if $\inf_{j,k\in\mathcal{I}, j\neq k} \|x_j - x_k\|_2 = \delta > 0$. Finally, by I_d we denote the identity operator on $\mathbb{R}^d$ and by χ_Q we denote the indicator function for a set $Q \subset \mathbb{R}^d$.

2 Decomposition spaces

In this section we define decomposition spaces [17] based on structured coverings [2]. For an invertible matrix $A \in GL(\mathbb{R}^d)$, and a constant $c \in \mathbb{R}^d$, we define the affine transformation $Tx = Ax + c$ with $x \in \mathbb{R}^d$. Given a family $\mathcal{T} = \{A_k(\cdot) + c_k\}_{k\in\mathbb{N}}$ of invertible affine transformations on $\mathbb{R}^d$, and a subset $Q \subset \mathbb{R}^d$, we let $\{Q_T\}_{T\in\mathcal{T}} := \{T(Q)\}_{T\in\mathcal{T}}$ and

$$\widetilde{T} := \left\{T' \in \mathcal{T} \;\middle|\; Q_{T'} \cap Q_T \neq \emptyset\right\}, \quad T \in \mathcal{T}. \tag{C.1}$$

We say that $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$ is an *admissible covering* of $\mathbb{R}^d$ if $\bigcup_{T\in\mathcal{T}} Q_T = \mathbb{R}^d$ and there exists $n_0 \in \mathbb{N}$ such that $|\widetilde{T}| \leq n_0$ for all $T \in \mathcal{T}$.

Definition 2.1 ($\mathcal{Q}-$moderate weight)**.** Let $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$ be an admissible covering. A function $u : \mathbb{R}^d \to (0,\infty)$ is called $\mathcal{Q}-$moderate if there exists $C > 0$ such that $u(x) \leq Cu(y)$ for all $x, y \in Q_T$ and all $T \in \mathcal{T}$. A $\mathcal{Q}-$moderate weight (derived from u) is a sequence $\{\omega_T\}_{T\in\mathcal{T}} := \{u(x_T)\}_{T\in\mathcal{T}}$ with $x_T \in Q_T$ for all $T \in \mathcal{T}$.

For the rest of this article, we shall use the explicit choice $u(x) := 1 + \|x\|_2$ for the function u in Definition 2.1. Let us now define structured coverings [2] of the time domain.

Definition 2.2 (Structured covering)**.** Given a family $\mathcal{T} = \{A_k(\cdot) + c_k\}_{k\in\mathbb{N}}$ of invertible affine transformations on $\mathbb{R}^d$, suppose there exist two bounded open sets $P \subset Q \subset \mathbb{R}^d$, with P compactly contained in Q, such that

1. $\{P_T\}_{T\in\mathcal{T}}$ and $\{Q_T\}_{T\in\mathcal{T}}$ are admissible coverings.
2. There exists a $\delta-$separated set $\{x_T\}_{T\in\mathcal{T}} \subset \mathbb{R}^d$, with $x_T \in Q_T$ for all $T \in \mathcal{T}$, such that $\{\omega_T\}_{T\in\mathcal{T}} := \{1 + \|x_T\|_2\}_{T\in\mathcal{T}}$ is a $\mathcal{Q}-$moderate weight.

Then we call $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ a *structured covering*.

For a structured covering we have the associated concept of a *bounded admissible partition of unity* (BAPU) [17].

Definition 2.3 (BAPU)**.** Let $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ be a structured covering of $\mathbb{R}^d$. A BAPU subordinate to $\mathcal{Q}$ is a family of non-negative functions $\{\psi_T\}_{T\in\mathcal{T}} \subset C_c^\infty(\mathbb{R}^d)$ satisfying

1. $\operatorname{supp}(\psi_T) \subset Q_T, \quad \forall T \in \mathcal{T}$.
2. $\sum\limits_{T\in\mathcal{T}} \psi_T(x) = 1, \quad \forall x \in \mathbb{R}^d$.

We note that the assumptions in Definition 2.3 imply that the members of the BAPU are uniformly bounded, i.e., $\sup_{T\in\mathcal{T}} \|\psi_T\|_{L^\infty} \leq 1$. Given a structured covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$, we can always construct a subordinate BAPU. Choose a non-negative function $\Phi \in C_c^\infty(\mathbb{R}^d)$, with $\Phi(x) = 1$ for all $x \in P$ and $\operatorname{supp}(\Phi) \subset Q$, and define

$$\psi_T(x) := \frac{\Phi(T^{-1}x)}{\sum_{T'\in\mathcal{T}} \Phi(T'^{-1}x)}, \quad x \in \mathbb{R}^d,$$

for all $T \in \mathcal{T}$. With this construction, it is clear that Definition 2.3(1) is satisfied. Furthermore, since $\{P_T\}_{T\in\mathcal{T}}$ is an admissible covering, then $1 \leq \sum_{T'\in\mathcal{T}} \Phi(T'^{-1}x) \leq n_0$ for all $x \in \mathbb{R}^d$ which shows that Definition 2.3(2) holds.

Remark 2.1. We note that the assumption in Definition 2.2(2) is not necessary for constructing a subordinate BAPU, however, the assumption is needed for proving Theorem 2.1.

Let $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ be a structured covering with $\mathcal{Q}-$moderate weight $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|x_T\|_2\}_{T\in\mathcal{T}}$ and BAPU $\{\psi_T\}_{T\in\mathcal{T}}$. For $s \in \mathbb{R}$ and $1 \leq q \leq \infty$, we define the associated weighted sequence space

$$\ell^q_{\omega^s}(\mathcal{T}) := \left\{ \{a_T\}_{T\in\mathcal{T}} \subset \mathbb{C} \;\middle|\; \|\{a_T\}_{T\in\mathcal{T}}\|_{\ell^q_{\omega^s}} := \|\{\omega_T^s a_T\}_{T\in\mathcal{T}}\|_{\ell^q} < \infty \right\}.$$

Given $\{a_T\}_{T\in\mathcal{T}} \in \ell^q_{\omega^s}(\mathcal{T})$, we define $\{a_T^+\}_{T\in\mathcal{T}}$ by $a_T^+ := \sum_{T'\in\widetilde{T}} a_{T'}$. Since $\{\omega_T\}_{T\in\mathcal{T}}$ is $\mathcal{Q}-$moderate, $\{a_T\}_{T\in\mathcal{T}} \to \{a_T^+\}_{T\in\mathcal{T}}$ defines a bounded operator on $\ell^q_{\omega^s}(\mathcal{T})$ according to [17, Remark 2.13 and Lemma 3.2]. Denoting its operator norm by C_+, we have

$$\left\|\{a_T^+\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}} \leq C_+ \left\|\{a_T\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}}, \quad \forall\, \{a_T\}_{T\in\mathcal{T}} \in \ell^q_{\omega^s}(\mathcal{T}). \tag{C.2}$$

We now define decomposition spaces as first introduced in [17].

Definition 2.4 (Decomposition space)**.** Let $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ be a structured covering with $\mathcal{Q}-$moderate weight $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|x_T\|_2\}_{T\in\mathcal{T}}$ and BAPU $\{\psi_T\}_{T\in\mathcal{T}}$. For $s \in \mathbb{R}$ and $1 \leq p, q \leq \infty$, we define the *decomposition space* $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ as the set of distributions $f \in \mathcal{S}'(\mathbb{R}^d)$ satisfying

$$\|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})} := \left\|\{\|\psi_T f\|_{L^p}\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}} < \infty.$$

Remark 2.2. According to [17, Theorem 3.7], $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ is independent of the particular choice of BAPU and different choices yield equivalent norms. Actually the results in [17] show that $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ is invariant under certain geometric modifications of $\mathcal{Q}$, but we will not go into detail here.

Remark 2.3. In contrast to the approach taken in [32] (where the decomposition is performed on the frequency side), we do not allow $p, q < 1$ in Definition 2.4 since a simple consideration shows that the resulting decomposition spaces would not be complete in this case.

We now consider some familiar examples of decomposition spaces. By standard arguments it is easy to verify that $D(\mathcal{Q}, L^2, \ell^2) = L^2(\mathbb{R}^d)$ with equivalent norms for any structured covering $\mathcal{Q}$. The next example shows how to construct Wiener amalgam spaces.

Example 2.1. Let $Q \subset \mathbb{R}^d$ be an open cube with center 0 and side-length $r > 1$. Define $\mathcal{T} := \{T_k\}_{k\in\mathbb{Z}^d}$, with $T_k x := x - k$ for all $k \in \mathbb{Z}^d$, and let $\{\omega_{T_k}\}_{T_k\in\mathcal{T}} = \{1 + \|k\|_2\}_{T_k\in\mathcal{T}}$. With $\mathcal{Q} := \{Q_{T_k}\}_{T_k\in\mathcal{T}}$, then $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ corresponds to the Wiener amalgam space $W(L^p, \ell^q_{\omega^s})$ for $s \in \mathbb{R}$ and $1 \leq p, q \leq \infty$, see [13] for further details. △

Let us now prove the following important properties of decomposition spaces.

Theorem 2.1. *Let $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ be a structured covering with $\mathcal{Q}-$moderate weight $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|x_T\|_2\}_{T\in\mathcal{T}}$ and subordinate BAPU $\{\psi_T\}_{T\in\mathcal{T}}$. For $s \in \mathbb{R}$ and $1 \le p, q \le \infty$,*

1. *$\mathcal{S}(\mathbb{R}^d) \hookrightarrow D(\mathcal{Q}, L^p, \ell^q_{\omega^s}) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$.*
2. *$D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ is a Banach space.*
3. *If $p, q < \infty$, then $\mathcal{S}(\mathbb{R}^d)$ is dense in $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$.*
4. *If $p, q < \infty$, then the dual space of $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ can be identified with $D(\mathcal{Q}, L^{p'}, \ell^{q'}_{\omega^{-s}})$ with $1/p + 1/p' = 1$ and $1/q + 1/q' = 1$.*

The proof of Theorem 2.1 can be found in Appendix A. In the next section we construct decomposition spaces, which are compatible with the structure of painless NSGFs with flexible time resolution.

3 Nonstationary Gabor frames

In this section, we construct NSGFs with flexible time resolution using the notation of [1]. Given a set of window functions $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$, with corresponding frequency sampling steps $b_n > 0$, then for $m, n \in \mathbb{Z}^d$ we define atoms of the form

$$g_{m,n}(x) := g_n(x)e^{2\pi i m b_n \cdot x}, \quad x \in \mathbb{R}^d.$$

The choice of $\mathbb{Z}^d$ as index set for n is only a matter of notational convenience; any countable index set would do.

Example 3.1. With $g_n(x) := g(x - na)$ and $b_n := b$ for all $n \in \mathbb{Z}^d$ we get

$$g_{m,n}(x) := g(x - na)e^{2\pi i m b \cdot x}, \quad x \in \mathbb{R}^d,$$

which just corresponds to a standard Gabor system. △

If $\sum_{m,n} |\langle f, g_{m,n}\rangle|^2 \asymp \|f\|_2^2$ for all $f \in L^2(\mathbb{R}^d)$, we refer to $\{g_{m,n}\}_{m,n}$ as an NSGF. For an NSGF $\{g_{m,n}\}_{m,n}$, the frame operator

$$Sf = \sum_{m,n\in\mathbb{Z}^d} \langle f, g_{m,n}\rangle\, g_{m,n}, \quad f \in L^2(\mathbb{R}^d),$$

is invertible and we have the expansions

$$f = \sum_{m,n\in\mathbb{Z}^d} \langle f, g_{m,n}\rangle\, \tilde{g}_{m,n}, \quad f \in L^2(\mathbb{R}^d),$$

with $\{\tilde{g}_{m,n}\}_{m,n} := \{S^{-1}g_{m,n}\}_{m,n}$ being the canonical dual frame of $\{g_{m,n}\}_{m,n}$ [5]. For notational convenience we define $G(x) := \sum_{n\in\mathbb{Z}^d} 1/b_n^d\, |g_n(x)|^2$. With this notation we have the following result [1, Theorem 1].

Theorem 3.1. *Let $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ with frequency sampling steps $\{b_n\}_{n\in\mathbb{Z}^d}$, $b_n > 0$ for all $n \in \mathbb{Z}^d$. Assuming supp$(g_n) \subseteq [0, \frac{1}{b_n}]^d + a_n$, with $a_n \in \mathbb{R}^d$ for all $n \in \mathbb{Z}^d$, the frame operator for the system*

$$g_{m,n}(x) = g_n(x)e^{2\pi i m b_n \cdot x}, \quad \forall m, n \in \mathbb{Z}^d, \quad x \in \mathbb{R}^d,$$

is given by

$$Sf(x) = G(x)f(x), \quad f \in L^2(\mathbb{R}^d).$$

The system $\{g_{m,n}\}_{m,n\in\mathbb{Z}^d}$ constitutes a frame for $L^2(\mathbb{R}^d)$, with frame-bounds $0 < A \leq B < \infty$, if and only if

$$A \leq G(x) \leq B, \quad \text{for a.e. } x \in \mathbb{R}^d, \tag{C.3}$$

and the canonical dual frame is then given by

$$\tilde{g}_{m,n}(x) = \frac{g_n(x)}{G(x)} e^{2\pi i m b_n \cdot x}, \quad x \in \mathbb{R}^d. \tag{C.4}$$

Remark 3.1. We note that the canonical dual frame in (C.4) posses the same structure as the original frame, which is a property not shared by general NSGFs. We also note that the canonical tight frame can be obtained by taking the square root of the denominator in (C.4).

Traditionally, an NSGF satisfying the assumptions of Theorem 3.1 is called a *painless* NSGF, referring to the fact that the frame operator is a simple multiplication operator. This terminology is adopted from the classical *painless nonorthogonal expansions* [6], which corresponds to the painless case for classical Gabor frames. By slight abuse of notation we use the term "painless" to denote the NSGFs satisfying Definition 3.1 below. In order to properly formulate this definition, we first need some preliminary notation which we adopt from [32].

Let $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ satisfy the assumptions in Theorem 3.1. Given $C_* > 0$ we denote by $\{I_n\}_{n\in\mathbb{Z}^d}$ the open cubes

$$I_n := \left(-\varepsilon_n, \frac{1}{b_n} + \varepsilon_n\right)^d + a_n, \quad \forall n \in \mathbb{Z}^d, \tag{C.5}$$

with $\varepsilon_n := C_*/b_n$ for all $n \in \mathbb{Z}^d$. We note that supp$(g_{m,n}) \subset I_n$ for all $m, n \in \mathbb{Z}^d$. For $n \in \mathbb{Z}^d$ we define

$$\widetilde{n} := \left\{ n' \in \mathbb{Z}^d \mid I_{n'} \cap I_n \neq \emptyset \right\},$$

using the notation of (C.1).

Definition 3.1 (Painless NSGF). Let $\{g_n\}_{n\in\mathbb{Z}^d} \subset L^2(\mathbb{R}^d)$ satisfy the assumptions in Theorem 3.1, and assume further that,

1. There exists $C_* > 0$ and $n_0 \in \mathbb{N}$, such that the open cubes $\{I_n\}_{n\in\mathbb{Z}^d}$, given in (C.5), satisfy $|\widetilde{n}| \leq n_0$ uniformly for all $n \in \mathbb{Z}^d$.

2. $\{a_n\}_{n\in\mathbb{Z}^d}$ is a $\delta-$separated set and $\{1 + \|a_n\|_2\}_{n\in\mathbb{Z}^d}$ constitutes a $\{I_n\}_{n\in\mathbb{Z}^d}-$moderate weight.

3. The g_n's are continuous, real valued and satisfy

$$g_n(x) \leq C b_n^{d/2} \chi_{I_n}(x), \quad \text{for all } n \in \mathbb{Z}^d,$$

   for some uniform constant $C > 0$.

Then we refer to $\{g_{m,n}\}_{m,n\in\mathbb{Z}^d}$ as a *painless* NSGF.

The assumptions in Definition 3.1 are easily satisfied, but the support condition in Theorem 3.1 is rather restrictive and implies a certain redundancy of the system. Nevertheless, we must assume some structure on the dual frame, which is not provided by general NSGFs. We choose the framework of painless NSGFs and base our arguments on the fact that the dual frame possess the same structure as the original frame. We expect it is possible to extend the theory developed in this article to a more general settings by imposing general existence results for NSGFs [12, 26, 39]. We now provide a simple example of a set of window functions satisfying Definition 3.1(3).

Example 3.2. Choose a continuous real valued function $\varphi \in L^2(\mathbb{R}^d) \setminus \{0\}$ with $\operatorname{supp}(\varphi) \subseteq [0,1]^d$. For $n \in \mathbb{Z}^d$ define

$$g_n(x) := b_n^{d/2} \varphi(b_n(x - a_n)), \quad x \in \mathbb{R}^d,$$

with $a_n \in \mathbb{R}^d$ and $b_n > 0$. Then $\operatorname{supp}(g_n) \subseteq [0, \frac{1}{b_n}] + a_n$ and Definition 3.1(3) is satisfied. $\triangle$

Following the approach taken in [32], we define $Q := (0,1)^d$ together with the set of affine transformations $\mathcal{T} := \{A_n(\cdot) + c_n\}_{n\in\mathbb{Z}^d}$ with

$$A_n := \left(2\varepsilon_n + \frac{1}{b_n}\right) \cdot I_d, \quad \text{and} \quad (c_n)_j := -\varepsilon_n + (a_n)_j, \quad 1 \leq j \leq d.$$

It is then easily shown that $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}} = \{I_n\}_{n\in\mathbb{Z}^d}$ forms a structured covering of $\mathbb{R}^d$ [32, Lemma 4.1]. Given $s \in \mathbb{R}$ and $1 \leq p, q \leq \infty$, we may therefore construct the associated decomposition space $D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ with $\{\omega_T\}_{T\in\mathcal{T}} := \{1 + \|a_n\|_2\}_{n\in\mathbb{Z}^d}$.

Example 3.3. Let $\{g_{m,n}\}_{m\in\mathbb{Z}^d,n\in\mathbb{Z}^d}$ be a painless NSGF according to Definition 3.1. Assume additionally that $K := \inf\{b_n\}_{n\in\mathbb{Z}^d} > 0$ and that Definition 3.1(1) and Definition 3.1(2) hold for the larger cubes $K_n := (-\varepsilon, 1/K+\varepsilon)^d + a_n$ for some $\varepsilon > 0$. Defining $Q := (0,1)^d$ and $\mathcal{T} := \{A_n(\cdot) + c_n\}_{n\in\mathbb{Z}^d}$, with

$$A_n := \left(2\varepsilon + \frac{1}{K}\right)\cdot I_d, \quad \text{and} \quad (c_n)_j := -\varepsilon + (a_n)_j, \quad 1 \le j \le d,$$

we obtain the structured covering $\mathcal{Q} := \{K_n\}_{n\in\mathbb{Z}^d}$. In this special case the associated decomposition space is the Wiener amalgam space $W(L^p, \ell^q_{\omega^s})$ for $s \in \mathbb{R}$ and $1 \le p, q \le \infty$ (cf. Example 2.1). $\triangle$

For the rest of this article, we write $\{g_{m,T}\}_{m\in\mathbb{Z}^d,T\in\mathcal{T}}$ for a painless NSGF with associated structured covering $\mathcal{Q} := \{Q_T\}_{T\in\mathcal{T}}$. With this notation, then $\operatorname{supp}(g_{m,T}) \subset Q_T$ for all $m \in \mathbb{Z}^d$ and all $T \in \mathcal{T}$. Similarly we write $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|a_T\|_2\}_{T\in\mathcal{T}}$ for the associated weight function.

4 Characterization of decomposition spaces

Using the notation of [2] we define the sequence space $d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})$ as the set of coefficients $\{c_{m,T}\}_{m\in\mathbb{Z}^d,T\in\mathcal{T}} \subset \mathbb{C}$ satisfying

$$\left\|\{c_{m,T}\}_{m\in\mathbb{Z}^d,T\in\mathcal{T}}\right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})} := \left\|\{\|\{c_{m,T}\}_{m\in\mathbb{Z}^d}\|_{\ell^p}\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}} < \infty,$$

for $s \in \mathbb{R}$ and $1 \le p, q \le \infty$. We can now prove the following important stability result.

Theorem 4.1. *Let $\{g_{m,T}\}_{m\in\mathbb{Z}^d,T\in\mathcal{T}}$ be a painless NSGF with associated structured covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ and weight function $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|a_T\|_2\}_{T\in\mathcal{T}}$. Fix $s \in \mathbb{R}$, $1 \le p \le 2$ and let $p' := p/(p-1)$. For $f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ and $1 \le q \le \infty$,*

$$\left\|\{\langle f, g_{m,T}\rangle\}_{m\in\mathbb{Z}^d,T\in\mathcal{T}}\right\|_{d(\mathcal{Q},\ell^{p'},\ell^q_{\omega^s})} \le C\,\|f\|_{D(\mathcal{Q},L^p,\ell^q_{\omega^s})}, \tag{C.6}$$

and for $h \in D(\mathcal{Q}, L^{p'}, \ell^q_{\omega^s})$ and $1 \le q < \infty$,

$$\|h\|_{D(\mathcal{Q},L^{p'},\ell^q_{\omega^s})} \le C' \left\|\{\langle h, g_{m,T}\rangle\}_{m\in\mathbb{Z}^d,T\in\mathcal{T}}\right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})}. \tag{C.7}$$

Proof. We first prove (C.6). Given $f \in D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$, since $\widetilde{\psi_T} := \sum_{T'\in\widetilde{T}} \psi_T \equiv 1$ on Q_T, then

$$\|\{\langle f, g_{m,T}\rangle\}_{m\in\mathbb{Z}^d}\|_{\ell^{p'}} = \left(\sum_{m\in\mathbb{Z}^d}\left|\left\langle \widetilde{\psi_T} f, g_{m,T}\right\rangle\right|^{p'}\right)^{1/p'}$$

$$= b_T^{-d/2}\left(\sum_{m\in\mathbb{Z}^d}\left|b_T^{d/2}\int_{\mathbb{R}^d}\widetilde{\psi_T}(x) f(x) g_T(x) e^{-2\pi i m b_T\cdot x}dx\right|^{p'}\right)^{1/p'},$$

with $b_T > 0$ being the frequency sampling step. Since $1 \leq p \leq 2$ we can use the Hausdorff-Young inequality [28, Theorem 2.1 on page 98], which together with Definition 3.1(3) imply

$$\left\| \{\langle f, g_{m,T} \rangle\}_{m \in \mathbb{Z}^d} \right\|_{\ell^{p'}} \leq b_T^{-d/2} \left\| \widetilde{\psi_T} f g_T \right\|_{L^p} \leq C_1 \left\| \widetilde{\psi_T} f \right\|_{L^p}.$$

Hence, using (C.2) we get

$$\begin{aligned} \left\| \{\langle f, g_{m,T} \rangle\}_{m \in \mathbb{Z}^d, T \in \mathcal{T}} \right\|_{d(\mathcal{Q}, \ell^{p'}, \ell^q_{\omega^s})} &\leq C_1 \left\| \left\{ \left\| \widetilde{\psi_T} f \right\|_{L^p} \right\}_{T \in \mathcal{T}} \right\|_{\ell^q_{\omega^s}} \\ &\leq C_2 \|f\|_{D(\mathcal{Q}, L^p, \ell^q_{\omega^s})}. \end{aligned}$$

Let us now prove (C.7). Given $h \in D(\mathcal{Q}, L^{p'}, \ell^q_{\omega^s})$ we may write the norm as

$$\|h\|_{D(\mathcal{Q}, L^{p'}, \ell^q_{\omega^s})} = \sup_{\sigma \in \mathcal{S}(\mathbb{R}^d), \|\sigma\|_{D(\mathcal{Q}, L^p, \ell^{q'}_{\omega^{-s}})} = 1} |\langle h, \sigma \rangle|, \qquad q' := q/(q-1), \quad \text{(C.8)}$$

since the dual space of $D(\mathcal{Q}, L^{p'}, \ell^q_{\omega^s})$ can be identified with $D(\mathcal{Q}, L^p, \ell^{q'}_{\omega^{-s}})$ and since $\mathcal{S}(\mathbb{R}^d)$ is dense in $D(\mathcal{Q}, L^p, \ell^{q'}_{\omega^{-s}})$. Given $\sigma \in \mathcal{S}(\mathbb{R}^d)$, with $\|\sigma\|_{D(\mathcal{Q}, L^p, \ell^{q'}_{\omega^{-s}})} = 1$, we write the frame expansion of σ with respect to $\{g_{m,T}\}_{m,T}$ and apply Hölder's inequality twice to obtain

$$\begin{aligned} |\langle h, \sigma \rangle| &\leq \sum_{T \in \mathcal{T}} \sum_{m \in \mathbb{Z}^d} |\langle \sigma, g_{m,T} \rangle \langle h, \tilde{g}_{m,T} \rangle| \\ &\leq \sum_{T \in \mathcal{T}} \left\| \{\langle \sigma, g_{m,T} \rangle\}_{m \in \mathbb{Z}^d} \right\|_{\ell^{p'}} \left\| \{\langle h, \tilde{g}_{m,T} \rangle\}_{m \in \mathbb{Z}^d} \right\|_{\ell^p} \\ &\leq \left\| \{\langle \sigma, g_{m,T} \rangle\}_{m \in \mathbb{Z}^d, T \in \mathcal{T}} \right\|_{d(\mathcal{Q}, \ell^{p'}, \ell^{q'}_{\omega^{-s}})} \left\| \{\langle h, \tilde{g}_{m,T} \rangle\}_{m \in \mathbb{Z}^d, T \in \mathcal{T}} \right\|_{d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})}. \end{aligned} \quad \text{(C.9)}$$

According to (C.6) then

$$\left\| \{\langle \sigma, g_{m,T} \rangle\}_{m,T} \right\|_{d(\mathcal{Q}, \ell^{p'}, \ell^{q'}_{\omega^{-s}})} \leq C_1 \|\sigma\|_{D(\mathcal{Q}, L^p, \ell^{q'}_{\omega^{-s}})} = C_1,$$

which combined with (C.9) and (C.3) yield

$$\begin{aligned} |\langle h, \sigma \rangle| &\leq C_1 \left\| \{\langle h, \tilde{g}_{m,T} \rangle\}_{m \in \mathbb{Z}^d, T \in \mathcal{T}} \right\|_{d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})} \\ &\leq C_2 \left\| \{\langle h, g_{m,T} \rangle\}_{m \in \mathbb{Z}^d, T \in \mathcal{T}} \right\|_{d(\mathcal{Q}, \ell^p, \ell^q_{\omega^s})}, \end{aligned} \quad \text{(C.10)}$$

with $C_2 := C_1/A$. Finally, combining (C.8) and (C.10) we arrive at

$$\|h\|_{D(\mathcal{Q},L^{p'},\ell^q_{\omega^s})} \le C_2 \left\| \{\langle h, g_{m,T}\rangle\}_{m\in\mathbb{Z}^d, T\in\mathcal{T}} \right\|_{d(\mathcal{Q},\ell^p,\ell^q_{\omega^s})},$$

which proves (C.7). □

We note that for $s \in \mathbb{R}$, $1 \le q < \infty$ and $p = 2$, Theorem 4.1 yields the equivalence

$$\|f\|_{D(\mathcal{Q},L^2,\ell^q_{\omega^s})} \asymp \left\| \{\langle f, g_{m,T}\rangle\}_{m\in\mathbb{Z}^d, T\in\mathcal{T}} \right\|_{d(\mathcal{Q},\ell^2,\ell^q_{\omega^s})}, \quad f \in D(\mathcal{Q}, L^2, \ell^q_{\omega^s}).$$

It follows that the *coefficient operator* $C : f \to \{\langle f, g_{m,T}\rangle\}_{m,T}$ is bounded from $D(\mathcal{Q}, L^2, \ell^q_{\omega^s})$ into $d(\mathcal{Q}, \ell^2, \ell^q_{\omega^s})$. We define the corresponding *reconstruction operator* as

$$R\left(\{c_{m,T}\}_{m\in\mathbb{Z}^d, T\in\mathcal{T}}\right) = \sum_{T\in\mathcal{T}} \sum_{m\in\mathbb{Z}^d} c_{m,T}\tilde{g}_{m,T}, \qquad \forall \{c_{m,T}\}_{m,T} \in d(\mathcal{Q}, \ell^2, \ell^q_{\omega^s}).$$

With this notation we have the following result.

Proposition 4.1. *Let $\{g_{m,T}\}_{m\in\mathbb{Z}^d, T\in\mathcal{T}}$ be a painless NSGF with associated structured covering $\mathcal{Q} = \{Q_T\}_{T\in\mathcal{T}}$ and weight function $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|a_T\|_2\}_{T\in\mathcal{T}}$. Given $s \in \mathbb{R}$ and $1 \le q < \infty$, the reconstruction operator R is bounded from $d(\mathcal{Q}, \ell^2, \ell^q_{\omega^s})$ onto $D(\mathcal{Q}, L^2, \ell^q_{\omega^s})$ and we have the expansions*

$$f = RC(f) = \sum_{m\in\mathbb{Z}^d, T\in\mathcal{T}} \langle f, g_{m,T}\rangle\, \tilde{g}_{m,T}, \quad f \in D(\mathcal{Q}, L^2, \ell^q_{\omega^s}), \tag{C.11}$$

with unconditional convergence.

Proof. We first prove that R is bounded. Given $\{c_{m,T}\}_{m,T} \in d(\mathcal{Q}, \ell^2, \ell^q_{\omega^s})$, (C.2) and (C.3) yield

$$\begin{aligned} \|R(\{c_{m,T}\}_{m,T})\|_{D(\mathcal{Q},L^2,\ell^q_{\omega^s})} &= \left\| \left\{ \left\| \psi_T \left(\sum_{T'\in\widetilde{T}} \sum_{m\in\mathbb{Z}^d} c_{m,T'}\tilde{g}_{m,T'} \right) \right\|_{L^2} \right\}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}} \\ &\le C_1 \left\| \left\{ \left\| \sum_{m\in\mathbb{Z}^d} c_{m,T} g_{m,T} \right\|_{L^2} \right\}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}}. \end{aligned} \tag{C.12}$$

Applying Definition 3.1(3) and the Hausdorff-Young inequality [28, Theorem 2.2 on page 99] we get

$$\left\| \sum_{m\in\mathbb{Z}^d} c_{m,T} g_{m,T} \right\|_{L^2}^2 \le C \int_{\mathbb{R}^d} \left| b_T^{d/2} \sum_{m\in\mathbb{Z}^d} c_{m,T} e^{2\pi i m b_T\cdot x} \right|^2 dx \le C \left\| \{c_{m,T}\}_{m\in\mathbb{Z}^d} \right\|_{\ell^2}^2. \tag{C.13}$$

Combining (C.12) and (C.13) we arrive at

$$\begin{aligned}\|R(\{c_{m,T}\}_{m,T})\|_{D(\mathcal{Q},L^2,\ell^q_{\omega^s})} &\leq C_2 \left\| \{ \|\{c_{m,T}\}_{m\in\mathbb{Z}^d}\|_{\ell^2} \}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}} \\ &= C_2 \left\|\{c_{m,T}\}_{m,T}\right\|_{d(\mathcal{Q},\ell^2,\ell^q_{\omega^s})}, \end{aligned} \tag{C.14}$$

which shows the boundedness of R. Let us now prove the unconditional convergence of (C.11). Given $f \in D(\mathcal{Q}, L^2, \ell^q_{\omega^s})$ we can find a sequence $\{f_k\}_{k\in\mathbb{N}} \subset \mathcal{S}(\mathbb{R}^d)$ such that $f_k \to f$ in $D(\mathcal{Q}, L^2, \ell^q_{\omega^s})$. For each k we have the expansion $f_k = RC(f_k)$ and by continuity of RC we get $f = RC(f)$. Given $\varepsilon > 0$, (C.14) implies that we can find a *finite* subset $F_0 \subseteq \mathbb{Z}^d \times \mathcal{T}$, such that for all finite sets $F \supseteq F_0$,

$$\left\| f - \sum_{(m,T)\in F} \langle f, g_{m,T}\rangle\, \tilde{g}_{m,T} \right\|_{D(\mathcal{Q},L^2,\ell^q_{\omega^s})} \leq C_2 \left\| \{\langle f, g_{m,T}\rangle\}_{(m,T)\notin F} \right\|_{d(\mathcal{Q},\ell^2,\ell^q_{\omega^s})} < \varepsilon.$$

According to [22, Proposition 5.3.1 on page 98], this property is equivalent to unconditional convergence. □

Based on Proposition 4.1, we can show some important properties of $\{g_{m,T}\}_{m\in\mathbb{Z}^d, T\in\mathcal{T}}$ in connection with nonlinear approximation theory [7, 8]. Assume $f \in D(\mathcal{Q}, L^2, \ell^2_{\omega^s})$, for $s \in \mathbb{R}$, and write the frame expansion

$$f = \sum_{m\in\mathbb{Z}^d, T\in\mathcal{T}} \langle f, g_{m,T}\rangle\, \tilde{g}_{m,T}. \tag{C.15}$$

Let $\{\theta_k\}_{k\in\mathbb{N}}$ be a rearrangement of the frame coefficients $\{\langle f, g_{m,T}\rangle\}_{m,T}$ such that $\{|\theta_k|\}_{k\in\mathbb{N}}$ constitutes a non-increasing sequence. Also, let f_N be the N-term approximation to f obtained by extracting the terms in (C.15) corresponding to the N largest coefficients $\{\theta_k\}_{k=1}^N$. Since R is bounded, [20, Theorem 6] implies that for each $1 \leq \tau < 2$,

$$\begin{aligned}\|f - f_N\|_{D(\mathcal{Q},L^2,\ell^2_{\omega^s})} &\leq C_1 \left\|\{\theta_k\}_{k>N}\right\|_{d(\mathcal{Q},\ell^2,\ell^2_{\omega^s})} \leq C_2 N^{-\alpha} \left\|\{\theta_k\}_{k\in\mathbb{N}}\right\|_{d(\mathcal{Q},\ell^\tau,\ell^\tau_{\omega^s})} \\ &= C_2 N^{-\alpha} \left\|\{\langle f, g_{m,T}\rangle\}_{k\in\mathbb{N}}\right\|_{d(\mathcal{Q},\ell^\tau,\ell^\tau_{\omega^s})}, \quad \alpha := 1/\tau - 1/2. \end{aligned} \tag{C.16}$$

We conclude that for $f \in D(\mathcal{Q}, L^2, \ell^2_{\omega^s})$, with frame coefficients in $d(\mathcal{Q}, \ell^\tau, \ell^\tau_{\omega^s})$, we obtain good approximations in $D(\mathcal{Q}, L^2, \ell^2_{\omega^s})$ by thresholding the frame coefficients in (C.15). The rate of the approximation is given by $\alpha \in (0, 1/2]$.

5 Numerical experiments

In this section we provide the numerical experiments, thresholding coefficients of both stationary and nonstationary Gabor expansions. We note that analysis with a stationary Gabor frame corresponds to analysis with the short-time Fourier transform (STFT) as the Gabor coefficients can be rewritten as

$$\langle f, g_{m,n} \rangle = \int_{\mathbb{R}^d} f(t)\overline{g(t-na)}e^{-2\pi imb\cdot t}dt = V_g f(na, mb), \quad f \in L^2(\mathbb{R}^d),$$

with $V_g f(na, mb)$ denoting the STFT of f, with respect to g, at time na and frequency mb.

For the implementation we use MATLAB 2017B and in particular we use the following two toolboxes: The LTFAT [34] (version 2.2.0 or above) available from `http://ltfat.github.io/` and the NSGToolbox [1] (version 0.1.0 or above) available from `http://nsg.sourceforge.net/`. The sound files we consider are part of the EBU-SQAM database [38], which consists of 70 test sounds sampled at 44.1 kHz. The test sounds form a large variety of speech and music including single instruments, classical orchestra, and pop music. Since music signals are continuous signals of finite energy, it make sense to consider them in the framework of decomposition spaces. Moreover, the decomposition space norm constitutes a natural measure for such nonstationary signals, capable of detecting local signal changes as opposed to the standard L^p-norm.

We divide the numerical analysis into two sections. In Section 5.1 we compare the performance of an adaptive nonstationary Gabor expansion to that of a classical Gabor expansion by analyzing spectrograms, reconstruction errors, and approximation rates associated to a particular music signal (signal 39 of the EBU-SQAM database). Then, in Section 5.2 we extend the experiment to cover the entire EBU-SQAM database and compare the average reconstruction errors and approximation rates, taken over the 70 test signals, for the two methods. To analyse the performance of an expansion we use the relative root mean square (RMS) reconstruction error

$$\mathrm{RMS}(f, f_{rec}) := \frac{\|f - f_{rec}\|_2}{\|f\|_2}.$$

As a general rule of thumb, an RMS error below 1% is hardly noticeable to the average listener. We measure the redundancy of a transform by

$$\frac{\text{number of coefficients}}{\text{length of signal}}.$$

The redundancy of the adaptive NSGF is approximately 5/3 and we have chosen parameters for the stationary Gabor frame, which mathes this redundancy.

5.1 Single experiment

In this experiment we consider sample 22000-284143 of signal 39 in the EBU-SQAM database. This signal is a piece of piano music consisting of an increasing melody of 10 individual tones (taken from an F major chord) starting at F2 (87 Hz fundamental frequency) and ending at F5 (698 Hz fundamental frequency). We construct the Gabor expansion using 1536 frequency channels and a hop size of 1024. The window function is chosen as a Hanning window of length 1536 such that the resulting system constitutes a painless Gabor frame. The Gabor transform has a redundancy of ≈ 1.51 and the total number of Gabor coefficients is 198402 (of which 195326 are non-zero). We only work with the coefficients of the positive frequencies since the signal is real valued. Performing hard thresholding, and keeping only the 15800 largest coefficients, we obtain a reconstructed signal with an RMS reconstruction error just below 1%.

For the adaptive NSGF, we choose to follow the adaptation procedure from [1], resulting in the construction of so-called *scale frames*. The idea is to calculate the onsets of the music piece, using a separate algorithm [9], and then to use short window functions around the onsets and long window functions between the onsets. The space between two onsets is spanned in such a way that the window length first increases (as we move away from the first onset) and then decreases (as we approach the second onset). To obtain a smooth resolution, the construction is such that adjacent windows are either of the same length or one is twice as long as the other. We refer the reader to [1] for further details. For the actual implementation, we use 8 different Hanning windows with lengths varying from 192 (around the onsets) to $192 \cdot 2^7 = 24576$. For the particular signal, the nonstationary Gabor transform has a redundancy of ≈ 1.66, which is comparable to that of the Gabor transform. The total number of coefficients is 217993 (of which 216067 are non-zero). Again, we only consider the coefficients of the positive frequencies. Keeping the 13100 largest coefficients we obtain an expansion with an RMS reconstruction error just below 1%. This is considerably fewer coefficients than needed for the stationary Gabor expansion, which shows a natural sparseness of scale frames for this particular signal class. This property was already noted by the authors in [1]. Spectrograms based on the original expansions and the thresholded expansions can be found in Fig. C.1.

The 10 "vertical stripes" in the spectrograms correspond to the onsets of the 10 tones in the melody and the "horizontal stripes" correspond to the frequencies of the harmonics. We note that the adaptive behaviour of the NSGF is clearly visible in the spectrograms, resulting in a good time resolution around the onsets and a good frequency resolution between the onsets. In contrast to this behaviour, the stationary Gabor frame uses a uniform res-

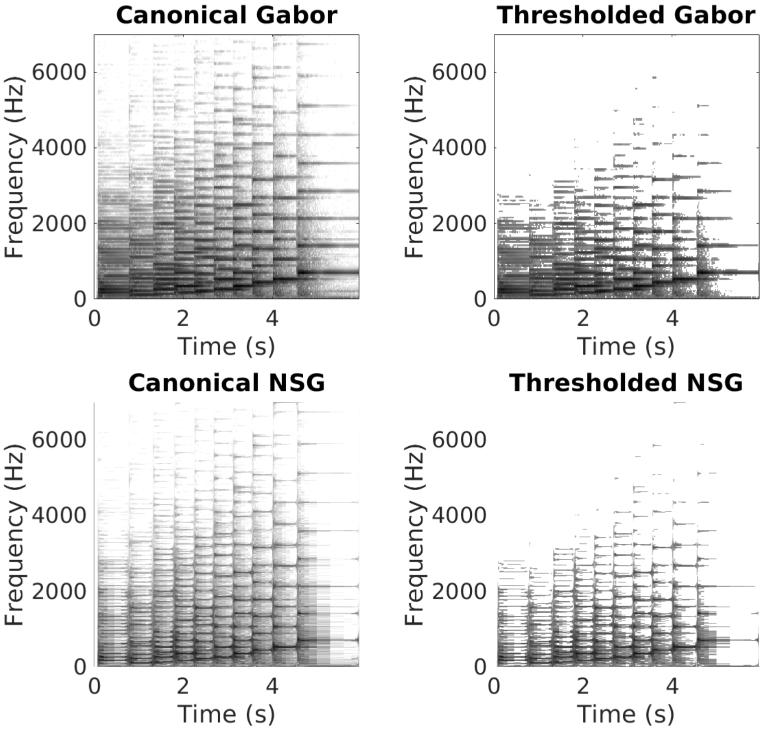

Fig. C.1: Spectrograms based on the original and thresholded Gabor- and nonstationary Gabor (NSG) expansions with RMS errors just below 1%.

olution over the whole time-frequency plane.

Based on the results from Section 4 (in particular (C.16)), we expect the RMS error $E(N)$ to decrease as $N^{-\alpha}$, for some $\alpha > 0$, with N being the number of non-zero coefficients. Calculating $E(N)$ for different values of N and performing power regression, we obtain the plots shown in Fig. C.2.

The results in Fig. C.2 show that both the RMS error $E(N)$ and the approximation rate α are lower for the nonstationary Gabor expansion than for the stationary Gabor expansion. Clearly, a small RMS error is more important than a fast approximation rate. Also, the fast approximation rate for the stationary Gabor frame is caused mainly by the high RMS error associated with small values of N. We note that both approximation rates are considerably faster than the rate given in (C.16) (which belongs to $(0, 1/2]$). This illustrates that (C.16) only provides us with an *upper bound* on the approximation error — the actual error might be much smaller. It also illustrates that both methods work extremely well for this kind of sparse signal. In the next section we

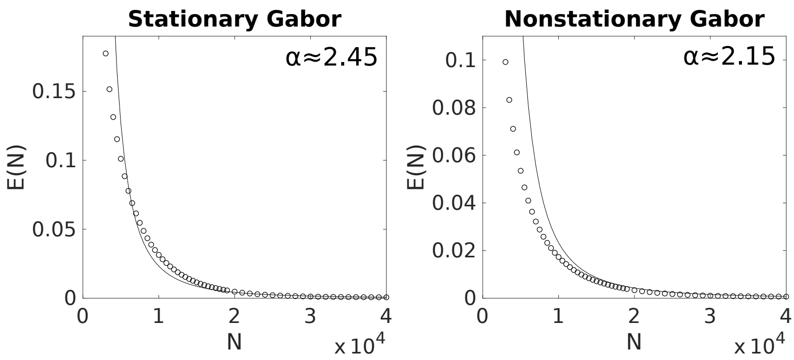

Fig. C.2: RMS error $E(N)$ as a function of N, the number of non-zero coefficients, for both stationary and nonstationary Gabor expansions. Also, an estimated power function is plotted for each expansion together with the associated value of the exponent α.

extend the analyzis presented here to cover the entire EBU-SQAM database.

5.2 Large scale experiment

For this experiment we consider the first 524288 samples of each of the 70 test sounds avaliable in the EBU-SQAM database. For each test sound we construct a nonstationary Gabor expansion, with parameters as described in Section 5.1, and three stationary Gabor expansions with different parameter settings. Using the notation (hopsize,number of frequency channels), we use the parameter settings $(1024, 2048)$, $(1536, 2048)$, and $(1024, 1536)$ for the three Gabor expansions. The window function associated to a Gabor expansion is chosen as a Hanning window with length equal to the corresponding number of frequency channels (resulting in a painless Gabor frame). For each of the four expansions we calculate for each test sound

1. The redundancy of the (non-thresholded) expansion.

2. Thresholded expansions with respect to N, the number of non-zero coefficient, where N takes on the values

$$N \in \{10000, 11000, \cdots, 29000, 30000, 35000, \cdots, 195000, 200000\}.$$

3. The sum of RMS errors $\sum_N E(N)$ taken over all the 55 possible values of N.

4. The value α of the estimated power function.

Repeating the experiment for all 70 test sounds we get the averaged values shown in Table C.1.

Table C.1: Average redundancies, sum of RMS errors, and approximation rates taken over the 70 test signals in the EBU-SQAM database. The experiment includes three stationary Gabor frames, with different parameters settings, and one NSGF.

Transform:	$G(1024, 2048)$	$G(1536, 2048)$	$G(1024, 1536)$	NSGF
Average redun.:	2.0020	1.3451	1.5049	1.6206
Average error:	2.1448	1.9128	1.9492	1.7367
Average α:	1.3088	1.4455	1.4278	1.2606

The results in Table C.1 show the same behaviour as the experiment in Section 5.1 — The NSGF provides the smallest RMS error and the slowest approximation rate. We note that the approximation rates all belong to the interval $[1.25; 1.45]$, which is much lower than the rates obtained in Section 5.1. This is due to the fact that the piano signal in Section 5.1 has a very sparse expansion, which is not true for all 70 test signals in the database. At first glance, the Gabor frame which seems to provide the best results is the one with parameter settings $(1536, 2048)$ — it produces the smallest RMS error and the largest approximation rate. However, this is mainly due to the low redundancy of the frame, which is only around 1.35. A low redundancy implies fewer Gabor coefficients (with more time-frequency information contained in each coefficient), which implies good results in terms of RMS error and approximation rate. However, a low redundancy also implies a worsened time-frequency resolution, which is not desirable for practical purposes. Finally, it is worth noting that the NSGF produces a significantly lower RMS error than the Gabor frame with parameters $(1536, 2048)$ even with a higher redundancy.

6 Conclusion

We have provided a self-contained description of decomposition spaces on the time side and proven several important properties of such spaces. Given a painless NSGF with flexible time resolution, we have shown how to construct an associated decomposition space, which characterizes signals with sparse expansions relative to the NSGF. Based on this characterization we have proven an upper bound on the approximation error occurring when thresholding the coefficients of the frame expansions. The theoretical results have been complemented with numerical experiments, illustrating that the approximation error is indeed smaller than the theoretical upper bound. Using terminology from nonlinear approximation theory, we have proven a Jackson inequality for nonlinear approximation with certain NSGFs. It could be interesting to consider the inverse estimate, a so-called Bernstein inequality, providing us with a lower bound on the approximation error. The numerical experiments indeed suggest that the approximation error acts as a power function of the number of non-zero coefficients. Unfortunately, obtaining a Bernstein inequality for such a redundant dictionary is in general beyond the reach of current methods [21].

A Proof of Theorem 2.1

Proof. We will use the well known fact that

$$\int_{\mathbb{R}^d} (1 + \|x\|_2)^{-m} dx < \infty, \quad m > d. \tag{C.17}$$

We prove each of the four statements separately and we write $D^s_{p,q} := D(\mathcal{Q}, L^p, \ell^q_{\omega^s})$ to simplify notation.

1. Repeating the arguments from [3, Proposition 5.7], using Definition 2.2(2), we can show that

$$D^{s+\varepsilon}_{p,\infty} \hookrightarrow D^s_{p,q} \hookrightarrow D^s_{p,\infty}, \quad \varepsilon > d/q, \tag{C.18}$$

for any $s \in \mathbb{R}$ and $1 \leq p, q \leq \infty$. Hence, to prove Theorem 2.1(1) it suffices to show that $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{p,\infty} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ for any $s \in \mathbb{R}$ and $1 \leq p \leq \infty$. We first show that $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{p,\infty}$. Since $\{\omega_T\}_{T\in\mathcal{T}} = \{1 + \|x_T\|_2\}_{T\in\mathcal{T}}$ is $\mathcal{Q}-$moderate, and ψ_T is uniformly bounded, this result follows from (C.17) since

$$\begin{aligned}
\omega_T^s \|\psi_T f\|_{L^p} &\leq C_1 \|(1 + \|\cdot\|_2)^s \psi_T f\|_{L^p} \leq C_1 \|(1 + \|\cdot\|_2)^s f\|_{L^p} \\
&\leq C_2 \left\|(1 + \|\cdot\|_2)^{s+r} f\right\|_{L^\infty} \\
&\leq C_2 \max_{|\beta|\leq N} \sup_{x\in\mathbb{R}^d} \left|(1 + \|x\|_2)^N \partial_x^\beta f(x)\right|, \quad f \in \mathcal{S}(\mathbb{R}^d),
\end{aligned}$$

for $r > d/p$ and $N \geq s + r$. To show that $D^s_{p,\infty} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$, we define $\widetilde{\psi_T} := \sum_{T'\in\widetilde{T}} \psi_{T'}$. Given $f \in D^s_{p,\infty}$ and $\varphi \in \mathcal{S}(\mathbb{R}^d)$, Hölder's inequality yields

$$\begin{aligned}
|\langle f, \varphi\rangle| &= \left|\sum_{T\in\mathcal{T}} \left\langle \psi_T f, \widetilde{\psi_T}\varphi\right\rangle\right| \leq \sum_{T\in\mathcal{T}} \left\|\psi_T f \widetilde{\psi_T}\varphi\right\|_{L^1} \\
&\leq \sum_{T\in\mathcal{T}} \|\psi_T f\|_{L^p} \left\|\widetilde{\psi_T}\varphi\right\|_{L^{p'}} \leq \|f\|_{D^s_{p,\infty}} \sum_{T\in\mathcal{T}} \omega_T^{-s} \left\|\widetilde{\psi_T}\varphi\right\|_{L^{p'}},
\end{aligned} \tag{C.19}$$

with $1/p + 1/p' = 1$. Applying (C.2) we get

$$\sum_{T\in\mathcal{T}} \omega_T^{-s} \left\| \widetilde{\psi_T}\varphi \right\|_{L^{p'}} \leq \left\| \left\{ \sum_{T'\in\widetilde{T}} \|\psi_{T'}\varphi\|_{L^{p'}} \right\}_{T\in\mathcal{T}} \right\|_{\ell^1_{\omega^{-s}}}$$
$$= \left\| \left\{ \left(\|\psi_T\varphi\|_{L^{p'}} \right)^+ \right\}_{T\in\mathcal{T}} \right\|_{\ell^1_{\omega^{-s}}}$$
$$\leq C_+ \left\| \left\{ \|\psi_T\varphi\|_{L^{p'}} \right\}_{T\in\mathcal{T}} \right\|_{\ell^1_{\omega^{-s}}} = C_+ \|\varphi\|_{D^{-s}_{p',1}}. \quad \text{(C.20)}$$

Now, (C.18) implies $\|\varphi\|_{D^{-s}_{p',1}} \leq C \|\varphi\|_{D^{\varepsilon-s}_{p',\infty}}$ for $\varepsilon > d$. Hence, since we have already shown that $\mathcal{S}(\mathbb{R}^d) \hookrightarrow D^s_{p,\infty}$, we conclude from (C.19) and (C.20) that $D^s_{p,\infty} \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$. This proves Theorem 2.1(1).

2. Theorem 2.1(2) follows from Theorem 2.1(1) and the arguments in [3, Page 150].

3. To prove Theorem 2.1(3) we let $f \in D^s_{p,q}$ and choose a function $I \in C_c^\infty(\mathbb{R}^d)$ satisfying $0 \leq I(x) \leq 1$ and $I(x) \equiv 1$ on some neighbourhood of $x = 0$. Since $\operatorname{supp}(I)$ is compact we can choose a *finite* subset $T^* \subset \mathcal{T}$ such that $\operatorname{supp}(I) \subset \cup_{T\in T^*} Q_T$ and $\sum_{T\in T^*} \psi_T(x) \equiv 1$ on $\operatorname{supp}(I)$. Hence, with $\widetilde{f} := If$ we get

$$\left\| \widetilde{f} \right\|_{L^p} = \left\| \sum_{T\in T^*} \psi_T I f \right\|_{L^p} \leq \sum_{T\in T^*} \|\psi_T f\|_{L^p} < \infty, \qquad \text{(C.21)}$$

since $f \in D^s_{p,q}$. Let $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $0 \leq \varphi(x) \leq 1$ and $\int_{\mathbb{R}^d} \varphi(x)dx = 1$. Also, for $\varepsilon > 0$ define $\varphi_\varepsilon(x) := \varepsilon^{-d}\varphi(x/\varepsilon)$ and let $\widetilde{f}_\varepsilon := \varphi_\varepsilon * \widetilde{f} \in \mathcal{S}(\mathbb{R}^d)$. It follows from (C.21) and a standard result on L^p-spaces [29, Theorem 2.16 on page 64] that

$$\left\| \widetilde{f} - \widetilde{f}_\varepsilon \right\|_{D^s_{p,q}} \leq \left\| \left\{ \left\| \widetilde{f} - \widetilde{f}_\varepsilon \right\|_{L^p} \right\}_{T\in\mathcal{T}} \right\|_{\ell^q_{\omega^s}} \to 0$$

as $\varepsilon \to 0$. Hence, the proof is done, if we can show that $\|f - \widetilde{f}\|_{D^s_{p,q}}$ can be made arbitrary small by choosing $\tilde{f}$ appropriately. To show this, we define $T_\circ := \{T \in \mathcal{T} \mid I(x) \equiv 1 \text{ on } \operatorname{supp}(\psi_T)\}$. Denoting its complement by $T_\circ^c$ we get

$$\left\| f - \widetilde{f} \right\|_{D^s_{p,q}} \leq 2 \left\| \{ \|\psi_T f\|_{L^p} \}_{T\in T_\circ^c} \right\|_{\ell^q_{\omega^s}}.$$

Finally, since $f \in D^s_{p,q}$, we can choose $\operatorname{supp}(I)$ large enough, such that $\|f - \widetilde{f}\|_{D^s_{p,q}} < \varepsilon$ for any given $\varepsilon > 0$. This proves Theorem 2.1(3).

4. To prove Theorem 2.1(4) we first note that $(D^s_{p,q})' \subset \mathcal{S}'(\mathbb{R}^d)$ since $\mathcal{S}(\mathbb{R}^d) \subset D^s_{p,q}$. Furthermore, by Remark 2.2 we may assume the same BAPU $\{\psi_T\}_{T\in\mathcal{T}}$ is used for both $D^s_{p,q}$ and $D^{-s}_{p',q'}$. Let us first show that $D^{-s}_{p',q'} \subseteq (D^s_{p,q})'$. Given $\sigma \in D^{-s}_{p',q'}$ and $f \in D^s_{p,q}$, applying (C.2) and Hölder's inequality twice yield

$$
\begin{aligned}
|\langle f,\sigma\rangle| &= \left|\sum_{T\in\mathcal{T}} \left\langle \widetilde{\psi_T} f, \psi_T\sigma\right\rangle\right| \le \sum_{T\in\mathcal{T}} \left\|\widetilde{\psi_T} f\right\|_{L^p} \|\psi_T\sigma\|_{L^{p'}} \\
&\le \sum_{T\in\mathcal{T}} \left(\omega_T^s \sum_{T'\in\widetilde{T}} \|\psi_{T'} f\|_{L^p}\right)\left(\omega_T^{-s}\|\psi_T\sigma\|_{L^{p'}}\right) \\
&\le \left\|\left\{\left\|(\psi_T f)^+\right\|_{L^p}\right\}_{T\in\mathcal{T}}\right\|_{\ell^q_{\omega^s}} \left\|\left\{\|\psi_T\sigma\|_{L^{p'}}\right\}_{T\in\mathcal{T}}\right\|_{\ell^{q'}_{\omega^{-s}}} \\
&\le C_+ \|f\|_{D^s_{p,q}} \|\sigma\|_{D^{-s}_{p',q'}}.
\end{aligned}
$$

To prove that $(D^s_{p,q})' \subseteq D^{-s}_{p',q'}$ we define the space $\ell^q(L^p)$ as those $\{f_T\}_{T\in\mathcal{T}} \subset \mathcal{S}'(\mathbb{R}^d)$ satisfying

$$
\|\{f_T\}_{T\in\mathcal{T}}\|_{\ell^q(L^p)} := \left\|\{\|f_T\|_{L^p}\}_{T\in\mathcal{T}}\right\|_{\ell^q} < \infty.
$$

With this notation we get

$$
\|f\|_{D^s_{p,q}} = \left\|\{\omega_T^s \|\psi_T f\|_{L^p}\}_{T\in\mathcal{T}}\right\|_{\ell^q} = \|\{\omega_T^s\psi_T f\}_{T\in\mathcal{T}}\|_{\ell^q(L^p)},
$$

for all $f \in D^s_{p,q}$. Since $f \to \{\omega_T^s\psi_T f\}_{T\in\mathcal{T}}$ defines an injective mapping from $D^s_{p,q}$ onto a subspace of $\ell^q(L^p)$, every $\sigma \in (D^s_{p,q})'$ can be interpreted as a functional on that subspace. By the Hahn-Banach theorem, σ can be extended to a continuous linear functional on $\ell^q(L^p)$ where the norm of σ is preserved. It thus follows from [37, Proposition 2.11.1 on page 177] that for $f \in D^s_{p,q}$ we may write

$$
\sigma(f) = \int_{\mathbb{R}^d} \sum_{T\in\mathcal{T}} \sigma_T(x)\omega_T^s\psi_T(x) f(x)dx, \quad \text{where} \tag{C.22}
$$

$$
\{\sigma_T(x)\}_{T\in\mathcal{T}} \in \ell^{q'}\left(L^{p'}\right), \quad \text{and} \quad \|\sigma\|_* = \|\{\sigma_T\}_{T\in\mathcal{T}}\|_{\ell^{q'}\left(L^{p'}\right)}, \tag{C.23}
$$

with $\|\sigma\|_* := \sup_{\|\{h_T\}\|_{\ell^q(L^p)}=1} |\sigma(\{h_T\})|$ denoting the standard norm on $(\ell^q(L^p))'$. From (C.22) we conclude that the proof is done if we can show that $\sum_{T\in\mathcal{T}} \sigma_T(x)\omega_T^s\psi_T(x) \in D^{-s}_{p',q'}$. This follows from (C.2) since

$$\begin{aligned}
\left\| \sum_{T\in\mathcal{T}} \sigma_T\omega_T^s\psi_T \right\|_{D^{-s}_{p',q'}} &= \left\| \left\{ \left\| \psi_T \left(\sum_{T'\in\widetilde{T}} \sigma_{T'}\omega_{T'}^s\psi_{T'} \right) \right\|_{L^{p'}} \right\}_{T\in\mathcal{T}} \right\|_{\ell^{q'}_{\omega^{-s}}} \\
&\leq C \left\| \left\{ \|\sigma_T\omega_T^s\psi_T\|_{L^{p'}} \right\}_{T\in\mathcal{T}} \right\|_{\ell^{q'}_{\omega^{-s}}} \\
&\leq C \left\| \left\{ \|\sigma_T\|_{L^{p'}} \right\}_{T\in\mathcal{T}} \right\|_{\ell^{q'}} \\
&= C \left\| \{\sigma_T\}_{T\in\mathcal{T}} \right\|_{\ell^{q'}(L^{p'})} = C\,\|\sigma\|_* ,
\end{aligned}$$

where we use (C.23) in the last equation. This proves Theorem 2.1(4).

□

References

[1] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco. Theory, implementation and applications of nonstationary Gabor frames. *J. Comput. Appl. Math.*, 236(6):1481–1496, 2011.

[2] L. Borup and M. Nielsen. Frame decomposition of decomposition spaces. *J. Fourier Anal. Appl.*, 13(1):39–70, 2007.

[3] L. Borup and M. Nielsen. On anisotropic Triebel-Lizorkin type spaces, with applications to the study of pseudo-differential operators. *J. Funct. Spaces Appl.*, 6(2):107–154, 2008.

[4] P. L. Butzer and K. Scherer. Jackson and Bernstein-type inequalities for families of commutative operators in Banach spaces. *J. Approximation Theory*, 5:308–342, 1972.

[5] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], second edition, 2016.

[6] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5):1271–1283, 1986.

[7] R. A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[8] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

[9] S. Dixon. Onset detection revisited. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 133–137, Montreal, Quebec, Canada, sep 2006.

[10] M. Dörfler. Time-frequency analysis for music signals: A mathematical approach. *Journal of New Music Research*, 30(1):3–12, 2001.

[11] M. Dörfler. Quilted Gabor frames—a new concept for adaptive time-frequency representation. *Adv. in Appl. Math.*, 47(4):668–687, 2011.

[12] M. Dörfler and E. Matusiak. Nonstationary Gabor frames—existence and construction. *Int. J. Wavelets Multiresolut. Inf. Process.*, 12(3):1450032, 18, 2014.

[13] H. G. Feichtinger. Banach convolution algebras of Wiener type. In *Functions, series, operators, Vol. I, II (Budapest, 1980)*, volume 35 of *Colloq. Math. Soc. János Bolyai*, pages 509–524. North-Holland, Amsterdam, 1983.

[14] H. G. Feichtinger. Modulation spaces on locally compact abelian groups. Technical report, University of Vienna, 1983.

[15] H. G. Feichtinger. Banach spaces of distributions defined by decomposition methods. II. *Math. Nachr.*, 132:207–237, 1987.

[16] H. G. Feichtinger. Modulation spaces: looking back and ahead. *Sampl. Theory Signal Image Process.*, 5(2):109–140, 2006.

[17] H. G. Feichtinger and P. Gröbner. Banach spaces of distributions defined by decomposition methods. I. *Math. Nachr.*, 123:97–120, 1985.

[18] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[19] D. Gabor. Theory of communication. *J. IEE*, 93(26):429–457, Nov. 1946.

[20] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10(1):51–71, 2004.

[21] R. Gribonval and M. Nielsen. Nonlinear approximation with dictionaries. II. Inverse estimates. *Constr. Approx.*, 24(2):157–173, 2006.

[22] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[23] K. Gröchenig and S. Samarah. Nonlinear approximation with local Fourier bases. *Constr. Approx.*, 16(3):317–331, 2000.

[24] C. Heil. An introduction to weighted Wiener amalgams. *Wavelets and their Applications (Chennai, January 2002)*, pages 183–216, 2003.

[25] E. Hernández, D. Labate, and G. Weiss. A unified characterization of reproducing systems generated by a finite family. II. *J. Geom. Anal.*, 12(4):615–662, 2002.

[26] N. Holighaus. Structure of nonstationary Gabor frames and their dual systems. *Appl. Comput. Harmon. Anal.*, 37(3):442–463, 2014.

[27] F. Jaillet and B. Torrésani. Time-frequency jigsaw puzzle: adaptive multi-window and multilayered Gabor expansions. *Int. J. Wavelets Multiresolut. Inf. Process.*, 5(2):293–315, 2007.

[28] Y. Katznelson. *An introduction to harmonic analysis*. Dover Publications, Inc., New York, corrected edition, 1976.

[29] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.

[30] M. Liuni, A. Röbel, E. Matusiak, M. Romito, and X. Rodet. Automatic adaptation of the time-frequency resolution for sound analysis and re-synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5):959–970, May 2013.

[31] S. Mallat. *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition, 2009. The sparse way, With contributions from Gabriel Peyré.

[32] E. S. Ottosen and M. Nielsen. A characterization of sparse nonstationary Gabor expansions. *Journal of Fourier Analysis and Applications*, May 2017.

[33] G. E. Pfander, H. Rauhut, and J. A. Tropp. The restricted isometry property for time-frequency structured random matrices. *Probab. Theory Related Fields*, 156(3-4):707–737, 2013.

[34] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs. The Large Time-Frequency Analysis Toolbox 2.0. In M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, editors, *Sound, Music, and Motion*, Lecture Notes in Computer Science, pages 419–442. Springer International Publishing, 2014.

[35] A. Ron and Z. Shen. Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbf{R}^d)$. *Duke Math. J.*, 89(2):237–282, 1997.

[36] A. Ron and Z. Shen. Generalized shift-invariant systems. *Constr. Approx.*, 22(1):1–45, 2005.

[37] H. Triebel. *Theory of function spaces*, volume 38 of *Mathematik und ihre Anwendungen in Physik und Technik [Mathematics and its applications in physics and technology]*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1983.

[38] E. B. Union. Sound quality assessment material: Recordings for subjective tests ; User's handbook for the EBU - SQAM compact disc, Sep 2008.

[39] F. Voigtlaender. Structured, compactly supported Banach frame decompositions of decomposition spaces. *ArXiv preprint*, arXiv:1612.08772, 2016.

[40] P. J. Wolfe, S. J. Godsill, and M. Dörfler. Multi-Gabor dictionaries for audio time-frequency analysis. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 43–46, 2001.

[41] R. M. Young. *An introduction to nonharmonic Fourier series*, volume 93 of *Pure and Applied Mathematics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.

[42] M. Zibulski and Y. Y. Zeevi. Analysis of multiwindow Gabor-type schemes by frame methods. *Appl. Comput. Harmon. Anal.*, 4(2):188–221, 1997.

Paper D

A Phase Vocoder Based on Nonstationary Gabor Frames

Emil Solsbæk Ottosen and Monika Dörfler

The paper has been published in
IEEE/ACM Transactions on Audio, Speech, and Language Processing Vol. 25(11), pp. 2199–2208, 2017.

The layout has been revised.

Abstract

We propose a new algorithm for time stretching music signals based on the theory of nonstationary Gabor frames (NSGFs). The algorithm extends the techniques of the classical phase vocoder (PV) by incorporating adaptive time-frequency (TF) representations and adaptive phase locking. The adaptive TF representations imply good time resolution for the onsets of attack transients and good frequency resolution for the sinusoidal components. We estimate the phase values only at peak channels and the remaining phases are then locked to the values of the peaks in an adaptive manner. During attack transients we keep the stretch factor equal to one and we propose a new strategy for determining which channels are relevant for reinitializing the corresponding phase values. In contrast to previously published algorithms we use a non-uniform NSGF to obtain a low redundancy of the corresponding TF representation. We show that with just three times as many TF coefficients as signal samples, artifacts such as phasiness and transient smearing can be greatly reduced compared to the classical PV. The proposed algorithm is tested on both synthetic and real world signals and compared with state of the art algorithms in a reproducible manner.

1 Introduction

The task of time stretching or pitch shifting music signals is fundamental in computer music and has many applications within areas such as transcription, mixing, transposition and auto-tuning [15, 25]. Time stretching is the operation of changing the length of a signal, without affecting its spectral content, while pitch shifting is the operation of raising or lowering the original pitch of a sound without affecting its length. As pitch shifting can be performed by combining time stretching and sampling rate conversion, we shall only focus on time stretching in this paper.

Introduced by Flanagan and Golden in [12], the phase vocoder (PV) stretches a signal by modifying its short time Fourier transform (STFT) in such a way that a stretched version can be obtain by reconstructing with respect to a different hop size. Through the years many improvements have been made and the PV is today a well-established technique [13, 17, 21, 22]. Unfortunately, it is known that the PV induces artifacts known as "phasiness" and "transient smearing" [17]. Phasiness is perceived as a characteristic colouration of the sound while transient smearing is heard as a lack of sharpness at the transients. Many modern techniques exist for dealing with these issues [10, 17, 26], but with only few exceptions [3, 7, 19], they are all based on the traditional idea of modifying a time-frequency (TF) representation obtained through the STFT. The STFT applies a sampling grid corresponding to a uniform TF resolution over the whole TF plane. For music signals it is often more appropriate to use good time resolution for the onset of attack transients and good frequency resolution for the sinusoidal components. We will consider the task of time stretching in the framework of Gabor theory [5, 14]. Applying nonstationary Gabor frames (NSGFs) [1, 9] we extend the theory of the PV to incorporate TF representations with the above-mentioned adaptive TF resolution.

In Section 1.1 of this article we describe some related work and explain the contributions of the proposed algorithm in relation to state of the art. In Section 2 we introduce the necessary tools from Gabor theory, including the painless condition for NSGFs. We use this framework to present the classical PV in Section 3 and the proposed algorithm in Section 4. We include the derivation of the classical PV for two reasons: Firstly, because it makes the transition to the nonstationary case easier and secondly, because we have not found any other thorough derivation in the literature that uses the framework of Gabor theory. Finally, in Section 5 we provide the numerical experiments and in Section 6 we give the conclusions.

1.1 State of the art

Traditionally, time-stretching algorithms are categorized into time-domain and frequency-domain techniques [21]. While time-domain techniques such as *synchronous overlap-add* (SOLA) [27] (and its extension PSOLA [4]) are capable of producing good results for monophonic signals, at a low computational cost, they tend to perform poorly when applied to polyphonic signals such as music. In contrast, frequency-domain methods, such as the PV [12], also work for polyphonic signals but with induced artifacts of their own, namely phasiness and transient smearing. As a first improvement to reduce phasiness, Puckette [24] suggested to use *phase-locking* to keep phase coherence intact over neighbouring frequency channels. This method was further studied by Laroche and Dolson [17] who proposed to separate the frequency axis into *regions of influences*, located around *peak channels*, and to lock the phase values of channels in a given region according to the phase value of the corresponding peak.

To deal with the issue of transient smearing, Bonada [3] proposed to keep the stretch factor equal to one during attack transients and then *reinitialize* all phase values for channels above a certain frequency cut, i.e. the phase values of these channels are set equal to the original phase values. In this way, the original timbre is kept intact without ruining the phase coherence for stationary partials at the lower frequencies. A more advanced approach for reducing transient smearing was presented by Röbel in [26]. Here, the transient detection algorithm works on the level of frequency channels and the reinitialization of a detected channel is performed for all time instants influenced by the transient. In this way, there is no need to set the stretch factor equal to one, which is a great advantage in regions with a dense set of transients.

More recent techniques have successfully reduced the PV artifacts by applying more sophisticated TF representations than the STFT. Bonada proposed the application of different FFTs for each time instant, which results in a TF representation with good frequency resolution at the lower frequencies and good time resolution at the higher frequencies. Derrien [7] suggested to construct an adaptive TF representation by choosing TF coefficients from a multi-scale Gabor dictionary under a matching constraint. A more recent algorithm, based on the theory of NSGFs, was proposed by Liuni et al. [19]. The idea behind their algorithm is to choose a fixed number of frequency bands and to apply, in each band, a NSGF with resolution varying in time. The window functions corresponding to the NSGFs are adapted to the signal by minimizing the *Rényi entropy*, which ensures a sparse TF representation. The techniques described in [26] and [19] are both implemented in the (commercialized) *super phase vocoder* (SuperVP) from IRCAM[1].

Contributions to state of the art

In order to generalize the techniques from the classical PV to the case where the TF representation is obtained through a NSGF, it is necessary to use the same number of frequency channels for each time instant. This construction corresponds to a *uniform* NSGF and, since the number of frequency channels must be at least equal to the length of the largest window function, necessarily leads to a high redundancy of the resulting transform.

In this paper we propose an algorithm, which fully exploits the potential of NSGFs to provide adaptivity while keeping a redundancy similar to the classical PV. This is achieved by letting the number of frequency channels for a given time instant equal the length of the window function selected for that particular time instant. This approach allows for using very long window functions, which is an advantage in regions with stationary partials. We summarize the contributions of this article as follows:

1. We explain the classical PV and the proposed algorithm in a unified framework using discrete Gabor theory.

2. We present a new time stretching algorithm, which uses an adaptive TF representation of lower redundancy than any other previously published algorithm.

3. While the proposed algorithm combines various familiar techniques from the literature, several new techniques are introduced in order to tackle the challenges arising from the application of non-uniform NSGFs. Hence, the proposed algorithm relies on techniques such as phase

[1] `http://anasynth.ircam.fr/home/english/software/supervp`

locking [17], transient detection [8], and quadratic interpolation [2] and integrates new methods for dealing with attack transients (cf. Section 4.2), for determining the phase values from frequencies estimated by quadratic interpolation (cf. Section 4.2), and for constructing the stretched signal from the modified (non-uniform) NSGF (cf. Section 4.3).

4. We provide a collection of sound files on-line (cf. Section 5) and include all source code necessary for reproducing the results.

2 Discrete Gabor theory

We write $f = (f[0], \dots, f[L-1])^T$ for a vector $f \in \mathbb{C}^L$ and $\mathbb{Z}_L = \{0, \dots, L-1\}$ for the cyclic group. Given $a, b \in \mathbb{Z}_L$, we define the *translation* operator $\mathbf{T}_a : \mathbb{C}^L \to \mathbb{C}^L$ and the *modulation* operator $\mathbf{M}_b : \mathbb{C}^L \to \mathbb{C}^L$ as

$$\mathbf{T}_a f[l] := f[l-a] \quad \text{and} \quad \mathbf{M}_b f[l] := f[l] e^{\frac{2\pi i b l}{L}},$$

for $l = 0, \dots, L-1$ and with translation performed modulo L. For $g \in \mathbb{C}^L$ and $a, b \in \mathbb{Z}_L$, we define the *Gabor system* $\{g_{m,n}\}_{m \in \mathbb{Z}_M, n \in \mathbb{Z}_N}$ as

$$g_{m,n}[l] := \mathbf{T}_{na} \mathbf{M}_{mb} g[l] = g[l - na] e^{\frac{2\pi i m b (l - na)}{L}},$$

with $Na = Mb = L$ for some $N, M \in \mathbb{N}$ [28, 29]. If $\{g_{m,n}\}_{m,n}$ spans $\mathbb{C}^L$, then it is called a *Gabor frame*. The associated *frame operator* $\mathbf{S} : \mathbb{C}^L \to \mathbb{C}^L$, defined by

$$\mathbf{S} f := \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle f, g_{m,n} \rangle \, g_{m,n}, \quad \forall f \in \mathbb{C}^L,$$

is invertible if and only if $\{g_{m,n}\}_{m,n}$ is a Gabor frame [5]. If $\mathbf{S}$ is invertible, then we have the expansions

$$f = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle f, g_{m,n} \rangle \, \tilde{g}_{m,n}, \quad \forall f \in \mathbb{C}^L, \tag{D.1}$$

with $\tilde{g}_{m,n} := \mathbf{T}_{na} \mathbf{M}_{mb} \mathbf{S}^{-1} g$. We say that $\{\tilde{g}_{m,n}\}_{m,n}$ is the *canonical dual frame* of $\{g_{m,n}\}_{m,n}$ and that $\mathbf{S}^{-1} g$ is the *canonical dual window* of g. The *discrete Gabor transform* (DGT) of $f \in \mathbb{C}^L$ is the matrix $\mathbf{c} \in \mathbb{C}^{M \times N}$ given by the coefficients $\{\langle f, g_{m,n} \rangle\}_{m,n}$ in the expansion (D.1). Finally, the ratio MN/L is called the *redundancy* of $\{g_{m,n}\}_{m,n}$.

Nonstationary Gabor frames

In this section we extend the classical Gabor theory to the nonstationary case [1]. Just as for the stationary case, we denote the total number of sampling points in time by $N \in \mathbb{N}$, however, we do not assume these points to be uniformly distributed. Further, instead of using just one window function, we apply $N_w \leq N$ different window functions $\{g_n\}_{n\in\mathbb{Z}_{N_w}}$ to obtain a flexible resolution. The window function corresponding to time point $n \in \mathbb{Z}_N$ is denoted by $g_{j(n)}$ with $j : \mathbb{Z}_N \to \mathbb{Z}_{N_w}$ being a surjective mapping. The number of frequency channels corresponding to time point $n \in \mathbb{Z}_N$ is denoted by $M_n \in \mathbb{Z}_L$ and the resulting frequency hop size by $b_n := L/M_n$. Finally, the window functions $\{g_n\}_{n\in\mathbb{Z}_{N_w}}$ are assumed to be symmetric around zero and we use translation parameters $\{a_n\}_{n\in\mathbb{Z}_N} \subset \mathbb{Z}_L$ to obtain to the proper support. With this notation, the *nonstationary Gabor system* (NSGS) $\{g_{m,n}\}_{m\in\mathbb{Z}_{M_n},n\in\mathbb{Z}_N}$ is defined as

$$g_{m,n}[l] := \mathbf{T}_{a_n}\mathbf{M}_{mb_n}g_{j(n)}[l] = g_{j(n)}[l-a_n]e^{\frac{2\pi imb_n(l-a_n)}{L}}.$$

If $\{g_{m,n}\}_{m,n}$ spans $\mathbb{C}^L$, then it is called a NSGF. If $M_n := M$, for all $n \in \mathbb{Z}_N$, then it is called a uniform NSGS (or uniform NSGF if it is also a frame). In Figure D.1 we see an example of a simple (non-uniform) NSGS with $N_w = 2$ and $N = 4$.

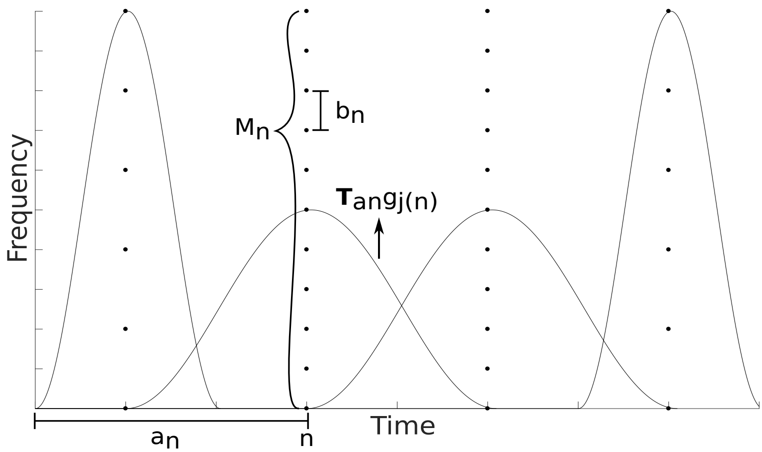

Fig. D.1: Illustration of a NSGS with $N_w = 2$ and $N = 4$.

Let us now show that the theory of NSGFs extends the theory of standard Gabor frames.

Example 2.1. Let $g \in \mathbb{C}^L$ and $a, b \in \mathbb{Z}_L$ satisfy $Na = Mb = L$ for some $N, M \in \mathbb{N}$. Then, with $g_{j(n)} := g$, $a_n := na$, and $b_n := b$ for all $n \in \mathbb{Z}_N$, we obtain the NSGS

$$g_{m,n}[l] = \mathbf{T}_{na}\mathbf{M}_{mb}g[l], \quad m \in \mathbb{Z}_M, \quad n \in \mathbb{Z}_N,$$

which just corresponds to a standard Gabor system. $\triangle$

The total number of elements in a NSGS $\{g_{m,n}\}_{m,n}$ is given by $P = \sum_{n=0}^{N-1} M_n$ and the redundancy is therefore P/L. The associated frame operator $\mathbf{S} : \mathbb{C}^L \to \mathbb{C}^L$, defined by

$$\mathbf{S}f := \sum_{n=0}^{N-1} \sum_{m=0}^{M_n-1} \langle f, g_{m,n} \rangle \, g_{m,n}, \quad \forall f \in \mathbb{C}^L,$$

is invertible if and only if $\{g_{m,n}\}_{m,n}$ constitutes a NSGF. If $\mathbf{S}$ is invertible, then we have the expansions

$$f = \sum_{n=0}^{N-1} \sum_{m=0}^{M_n-1} \langle f, g_{m,n} \rangle \, \tilde{g}_{m,n}, \quad \forall f \in \mathbb{C}^L, \tag{D.2}$$

with $\{\tilde{g}_{m,n}\}_{m,n} := \{\mathbf{S}^{-1}g_{m,n}\}_{m,n}$ being the canonical dual frame of $\{g_{m,n}\}_{m,n}$. The *nonstationary Gabor transform* (NSGT) of $f \in \mathbb{C}^L$ is given by the coefficients $\{c\{n\}(m)\}_{m,n} := \{\langle f, g_{m,n} \rangle\}_{m,n}$ in the expansion (D.2). We note that these coefficients do not form a matrix in the general case. We now consider an important case for which the calculation of $\{\tilde{g}_{m,n}\}_{m,n}$ is particularly simple.

Painless NSGFs

If $\mathrm{supp}(g_{j(n)}) \subseteq [c_{j(n)}, d_{j(n)}]$ and $d_{j(n)} - c_{j(n)} \leq M_n$ for all $n \in \mathbb{Z}_n$, then $\{g_{m,n}\}_{m,n}$ is called a *painless* NSGS (or painless NSGF if it is also a frame). In this case we have the following result [1].

Proposition 2.1. *If $\{g_{m,n}\}_{m,n}$ is a painless NSGS, then the frame operator $\mathbf{S}$ is an $L \times L$ diagonal matrix with entries*

$$S_{ll} = \sum_{n=0}^{N-1} M_n \left| g_{j(n)}[l - a_n] \right|^2, \quad \forall l \in \mathbb{Z}_L.$$

The system $\{g_{m,n}\}_{m,n}$ is a frame for $\mathbb{C}^L$ if and only if $\sum_{n=0}^{N-1} M_n \left| g_{j(n)}[l - a_n] \right|^2 > 0$ for all $l \in \mathbb{Z}_L$, and in this case the canonical dual frame $\{\tilde{g}_{m,n}\}_{m,n}$ is given by

$$\tilde{g}_{m,n}[l] = \frac{g_{m,n}[l]}{\sum_{n'=0}^{N-1} M_{n'} \left| g_{j(n')}[l - a_{n'}] \right|^2},$$

for all $n \in \mathbb{Z}_N$ and all $m \in \mathbb{Z}_{M_n}$.

We note that the canonical dual frame is also a painless NSGF, which is a property not shared by general NSGFs. An immediate consequence of Proposition 2.1 is the classical result for painless nonorthogonal expansions [6], which just corresponds to the painless case for standard Gabor frames.

3 The phase vocoder

In this section we explain the classical PV [17] in the framework of Gabor theory. The PV stretches the length of a signal by means of modifying its discrete STFT. Since the discrete STFT corresponds to a DGT, this technique can be perfectly well explained using Gabor theory. The main idea is to construct a DGT of the signal with respect to an *analysis* hop size a, modifying the DGT and then reconstructing from the modified DGT using a different *synthesis* hop size a_*. We only consider the case $a_* = ra$, for a constant modification rate $r > 0$. The case $r > 1$ corresponds to slowing down the signal by extending its length, while $r < 1$ corresponds to speeding it up by shortening its length. The PV is a classic analysis-modification-synthesis technique, and we will explain each of these three steps separately in the following sections.

3.1 Analysis

Let $\{g_{m,n}\}_{m,n}$ be a painless Gabor frame for $\mathbb{C}^L$. Given a real valued signal $f \in \mathbb{R}^L$, we calculate the DGT $\mathbf{c} \in \mathbb{C}^{M\times N}$ of f with respect to $\{g_{m,n}\}_{m,n}$ as

$$c_{m,n} = \langle f, g_{m,n}\rangle = \sum_{l=0}^{L-1} f[l]\overline{g[l-na]}e^{\frac{-2\pi imb(l-na)}{L}}, \tag{D.3}$$

for all $m \in \mathbb{Z}_M$ and $n \in \mathbb{Z}_N$. Let us explain the consequences of the phase convention used in (D.3). Define $\Omega_m := 2\pi m/M$ as the center frequency of the m'th channel and assume that g is real and symmetric around zero. Then, since $\{g_{m,n}\}_{m,n}$ is painless and $b/L = 1/M$, we may write (D.3) as

$$c_{m,n} = \sum_{l=0}^{M-1} f[l]g[na-l]e^{-i\Omega_m(l-na)} = e^{i\Omega_m na}\left(f_m * g\right)[na], \tag{D.4}$$

with $f_m[l] := f[l]e^{-i\Omega_m l}$. If g and $\hat{g}$ are both well-localized around zero, the convolution in (D.4) extracts the *baseband* spectrum of f_m at time na. Recalling that f_m is just a version of f that has been modulated down by m, this baseband spectrum corresponds to the spectrum of f in a neighbourhood of frequency m at time na. Finally, modulating back by m we obtain the *band-pass* spectrum of f in a neighbourhood of frequency m at time na. This phase convention is the traditional one used in the PV [17, 18, 21].

3.2 Modification

To explain the modification step of the PV, we refer to a quasi-stationary sinusoidal model that f is assumed to satisfy [16, 20]. This model is not used explicitly anywhere in the derivation of the PV, but it serves an important role for explaining the underlying ideas. We assume that f can be written as a *finite* sum of sinusoids

$$f(t) = \sum_k A_k(t) e^{i\theta_k(t)}, \tag{D.5}$$

in which $A_k(t)$ is the *amplitude*, $\theta_k(t)$ is the *phase*, and $\theta'_k(t)$ is the *frequency* of the k'th sinusoid at time t. Since the model is quasi-stationary, $A_k(t)$ and $\theta'_k(t)$ are assumed to be slowly varying functions. In particular, they are assumed to be almost constant over the duration of g. Based on (D.5), the perfectly stretched signal f_* at time $na_* = nra$ is given by

$$f_*[na_*] = \sum_k A_k(na) e^{ir\theta_k(na)}. \tag{D.6}$$

We note that the amplitudes and frequencies of the stretched signal f_* at time na_* equal the amplitudes and frequencies of the original signal f at time na.

The idea behind the modification step is to construct a new DGT $\mathbf{d} \in \mathbb{C}^{M\times N}$, based on $\mathbf{c} \in \mathbb{C}^{M\times N}$, such that reconstruction from $\mathbf{d}$, with respect to a_*, yields a time stretched version of f in the sense of (D.6). Since the amplitudes need to be preserved we set

$$d_{m,n} = |c_{m,n}|\, e^{i\angle d_{m,n}}, \quad m \in \mathbb{Z}_M, \quad n \in \mathbb{Z}_N,$$

using polar coordinates. Estimating the phases $\{\angle d_{m,n}\}_{m,n}$ involves a task called *phase unwrapping* [17].

Phase unwrapping

Assume there is a sinusoid of frequency ω in the vicinity of channel m at time na. Then, we make the estimate

$$e^{i\angle d_{m,n}} = e^{i(\angle d_{m,n-1} + \omega a_*)}, \tag{D.7}$$

since the two DGT samples $d_{m,n-1}$ and $d_{m,n}$ are a_* time samples apart. Using the same argument we may write $e^{i\angle c_{m,n}} = e^{i(\angle c_{m,n-1}+\omega a)}$. Setting $\omega = \Delta\omega + \Omega_m$, and isolating the deviation $\Delta\omega$, yields

$$\text{princarg}\left\{\Delta\omega a\right\} = \text{princarg}\left\{\angle c_{m,n} - \angle c_{m,n-1} - \Omega_m a\right\},$$

with "princarg" denoting the principal argument in the interval $]-\pi,\pi]$. Assuming ω is close to the center frequency Ω_m, such that $\Delta\omega \in]-\pi/a, \pi/a]$, we arrive at

$$\Delta\omega = \frac{\text{princarg}\left\{\angle c_{m,n} - \angle c_{m,n-1} - \Omega_m a\right\}}{a}.$$

We can now calculate ω as $\Delta\omega + \Omega_m$ and use (D.7) to determine $\{\angle d_{m,n}\}_{m,n}$ by initializing $d_{m,0} = c_{m,0}$ for all $m \in \mathbb{Z}_M$.

3.3 Synthesis

The final step of the PV is to construct a time stretched version of f in the sense of (D.6) from the modified DGT $\mathbf{d} \in \mathbb{C}^{M \times N}$. This is done by reconstructing from $\mathbf{d}$ with respect to the synthesis hop size a_*. According to (D.1), such a reconstruction yields

$$f_*[l] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} d_{m,n} \mathbf{T}_{na_*} \mathbf{M}_{mb} \mathbf{S}_*^{-1} g[l], \tag{D.8}$$

with $\mathbf{S}_* : \mathbb{C}^L \to \mathbb{C}^L$ being the modified frame operator

$$\mathbf{S}_* x[l] = \sum_{m=0}^{M-1} \sum_{n=0}^{N_*-1} \langle x, \mathbf{T}_{na_*} \mathbf{M}_{mb} g \rangle \, \mathbf{T}_{na_*} \mathbf{M}_{mb} g[l],$$

where $N_* := L/a_*$. The length of the reconstructed signal f_* is given by $L_* = Na_* = Lr$ and translation is performed modulo L_* in (D.8). In practice, the reconstruction formula (D.8) is realized by applying an inverse FFT and overlap-add.

Traditionally, a DGT with 75% overlap is used in the analysis step, which allows for modification factors $r \leq 4$. We note that if no modifications are made ($r = 1$), we recover the original signal. In the next section we consider some of the problems connected with the PV.

3.4 Drawbacks

The idea behind the PV is intuitive and easily implementable, which makes it attractive from a practical point of view. Unfortunately, the assumptions made in the modification part are not easily satisfied. This is true even for signals constructed explicitly from the sinusoidal model (D.5). We now list three main issues to be considered.

1. **Vertical coherence:** The PV ensures *horizontal coherence* [17] within each frequency channel but no attempt is made to ensure *vertical coherence* [17] across the frequency channels. If a sinusoid moves from one channel to another, the corresponding phase estimate might change dramatically. This is undesirable since a small change in frequency should only imply a small change in phase.

2. **Resolution:** In practice, we cannot assume that the sinusoids constituting f are well resolved in the DGT in the sense at most one sinusoid is

present within each frequency channel. The channels will only provide a "blurred" image of various neighbouring sinusoids. Furthermore, the amplitudes and frequencies of each sinusoid will often not be constant over the entire duration of g. As a consequence, the estimates made in the modification part will be subject to error.

3. **Transients:** The presence of attack transients is not well modelled within the PV as the phase values at such time instants cannot be predicted from previous estimates. Also, for music signals we often want the onsets to stay intact after time stretching, which is not accounted for in the PV approach.

In the next section we construct a new PV, which addresses the problems mentioned above.

4 A phase vocoder based on nonstationary Gabor frames

As mentioned in the introduction, the DGT is not always preferable for representing music signals as it corresponds to a uniform resolution over the whole TF plane. A poor TF resolution conflicts with the fundamental idea of well resolved sinusoids and therefore causes problems for the PV. In this section we change the TF representation from the DGT to an adaptable NSGT, which better matches the sinusoidal model (D.5). To be consistent with the description of the PV in Section 3 we separately explain the analysis, modification, and synthesis steps of the proposed algorithm.

4.1 Analysis

First of all, an adaptation procedure must be chosen for the NSGT. We choose to work with the procedure described in [1] since it is suitable for representing signals, which consist mainly of transient and sinusoidal components. The adaptation procedure is based on the idea that window functions with small support should be used around the onsets of attack transients, while window functions with longer support should be used between these onsets.

Remark 4.1. The construction presented here necessarily yields the problem of a coarse frequency resolution for the transient regions. However, as we propose to keep the stretch factor equal to one during attack transients (cf. Section 4.2), the impact of this problem is limited.

The onsets are calculated using a separate algorithm [8] and the window functions are constructed as scaled versions of a single window prototype (a Hanning window or similar). The resulting system is referred to as a *scale frame*. In the following paragraphs we explain the construction of scale frames in details.

Transient detection

To perform the transient detection we use a spectral flux (SF) onset function as described in [1, 8]. This function is computed with a DGT of redundancy 16, and it measures the sum of (positive) change in magnitude for all frequency channels. A time instant, corresponding to a local maximum of the SF function, is determined as an onset if its SF value is larger than the SF mean value in a certain neighbourhood of time frames. Hence, for region with a dense set of transients, only the most significant onsets are calculated. It is clear that such an approach must be taken to avoid an undesirably low frequency resolution in such regions. The redundant DGT used for the SF onset function is not used anywhere else in our algorithm and does not contribute significantly to the overall complexity.

Constructing the window functions

After a set of onsets has been extracted, the window functions are constructed following the rule that the space between two onsets is spanned in such a way that the window length first increases (as we get further away from the first onset) and then decreases (as we approach the next onset). The construction is performed in a smooth way such that the change from one step to the next corresponds to a window function that is either half as long, twice as long or of the same length. For details see [1]. The overlap between the window functions is chosen such that at most one onset is present within each time frame, we shall elaborate further on this particular construction in Section 4.3.

Constructing the NSGT

Once the window functions $\{g_n\}_{n\in\mathbb{Z}_{N_w}}$ have been constructed, we choose the numbers of frequency channels $\{M_n\}_{n\in\mathbb{Z}_N}$ such that the resulting system constitutes a painless NSGF. Additionally, we choose a lower bound on $\{M_n\}_{n\in\mathbb{Z}_N}$ to avoid an undesirably low number of channels around the onsets (explicit choices of parameters are described in Section 6). Given a real valued signal $f \in \mathbb{R}^L$, we calculate the NSGT $\{c\{n\}(m)\}_{m\in\mathbb{Z}_{M_n},n\in\mathbb{Z}_N}$ of f with respect to the scale frame $\{g_{m,n}\}_{m,n}$ as

$$c\{n\}(m) = \langle f, g_{m,n}\rangle = \sum_{l=0}^{L-1} f[l]\overline{g_{j(n)}[l-a_n]}e^{\frac{-2\pi imb_n(l-a_n)}{L}},$$

for all $n \in \mathbb{Z}_N$ and all $m \in \mathbb{Z}_{M_n}$. We note that the phase convention is the same as used in the PV (cf. Section 3.1).

4.2 Modification

The idea behind the modification step is the same as for the PV. We assume f satisfies (D.5), and we construct a modified NSGT $\{d\{n\}(m)\}_{m,n}$, based on $\{c\{n\}(m)\}_{m,n}$, such that reconstruction from $\{d\{n\}(m)\}_{m,n}$, with respect to a set of synthesis translation parameters, yields a time stretched version of f in the sense of (D.6). Given a stretch factor $r > 0$, the distance between synthesis time sample n and $n+1$ is

$$a_n^* := r(a_{n+1} - a_n), \quad n \in \mathbb{Z}_N. \tag{D.9}$$

Since we do not want the transients to be stretched, we let $r = 1$, when a_n corresponds to the onset of a transient, and then stretch with a correspondingly larger factor $r' > r$ in remaining regions. Using polar coordinates we set

$$d\{n\}(m) = |c\{n\}(m)|\, e^{i\angle d\{n\}(m)}, \quad n \in \mathbb{Z}_N, \quad m \in \mathbb{Z}_{M_n},$$

with $\angle d\{0\}(m) = \angle c\{0\}(m)$ for all $m \in \mathbb{Z}_{M_0}$. Hence, in complete analogy with the approach in the PV, the problem boils down to estimating the phase values $\{\angle d\{n\}(m)\}_{m,n}$.

Making the transition from stationary Gabor frames to NSGFs, we are facing a fundamental problem. The DGT corresponds to a uniform sampling grid over the TF plane, while the NSGF corresponds to a sampling grid which is irregular over time but regular over frequency for each fixed time position. This is illustrated in Figure D.2.

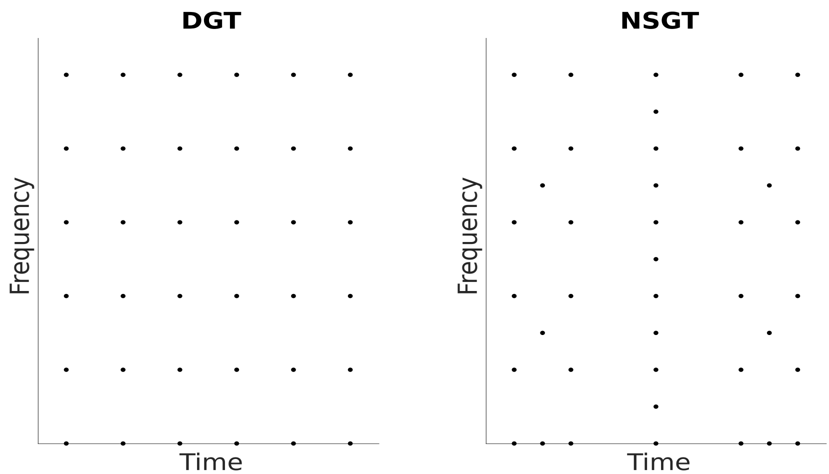

Fig. D.2: Sampling grids corresponding to a DGT and a NSGT.

As a consequence, we cannot guarantee that each sampling point has a horizontal neighbour that can be used for estimating the frequency as in the

PV (cf. Section 3). We therefore generalize the approach from [2] to the nonstationary case and calculate the frequencies using *quadratic interpolation*.

Calculating the frequencies

For fixed $n \in \mathbb{Z}_N$, we define channel m_p as a *peak* if its magnitude $|c\{n\}(m_p)|$ is larger than the magnitudes of its two vertical neighbors, i.e. $|c\{n\}(m_p)| > |c\{n\}(m_p \pm 1)|$. If there is a sinusoid of frequency ω in the vicinity of peak channel m_p, the "true" peak position will differ from m_p unless ω is exactly equal to $2\pi m_p / M_n$. The idea is thus to interpolate the true peak position, using the neighboring channels $m_p \pm 1$, and then to apply this value as an estimate for ω. To describe the setup we set the position of the peak channel m_p to 0, and the positions of its two neighbors to -1 and 1, respectively. Also, we denote the true peak position by p and define

$$\alpha := |c\{n\}(-1)|, \quad \beta := |c\{n\}(0)|, \quad \text{and} \quad \gamma := |c\{n\}(1)|.$$

The situation is illustrated in Figure D.3, with y denoting the parabola to be interpolated.

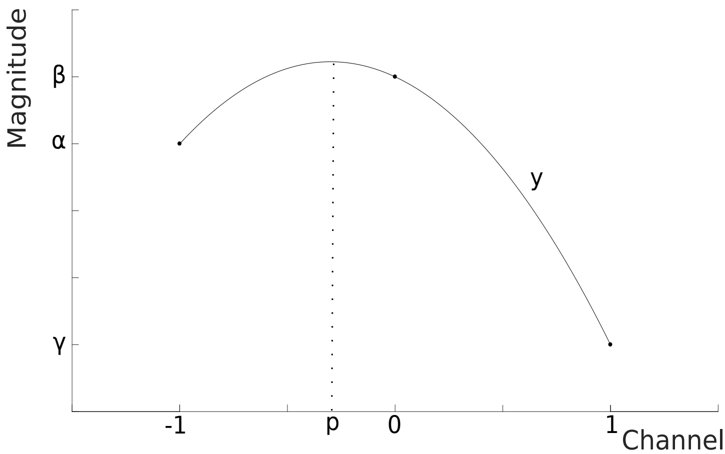

Fig. D.3: Illustration of quadratic interpolation.

Writing $y(x) = a(x-p)^2 + b$ and solving for p yields

$$p = \frac{1}{2} \cdot \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \in \left(-\frac{1}{2}, \frac{1}{2}\right).$$

The value of p determines the deviation from the peak channel to the true peak proportional to the size of the channel. After p has been determined, we calculate the frequency as

$$\omega = \frac{2\pi(m_p + p)}{M_n}. \tag{D.10}$$

In practice, the calculations are done on a dB scale for higher accuracy of the quadratic interpolation. Let us now explain how the frequency estimate (D.10) is used to calculate the corresponding phase value $\angle d\{n\}(m_p)$.

Calculating the phases

Between each pair of peaks we define the (lowest) channel with smallest magnitude as a *valley* and then use these valleys to separate the frequency axis into *regions of influence*. As noted in [17], if a peak switches from channel $m_{p'}$ at time $n-1$ to channel m_p at time n, the corresponding phase estimate should take this behavior into account. A simple way of determining the previous peak $m_{p'}$ is to choose the peak of the corresponding region of influence that channel m_p would have belonged to in time frame $n-1$. This is illustrated in Figure D.4.

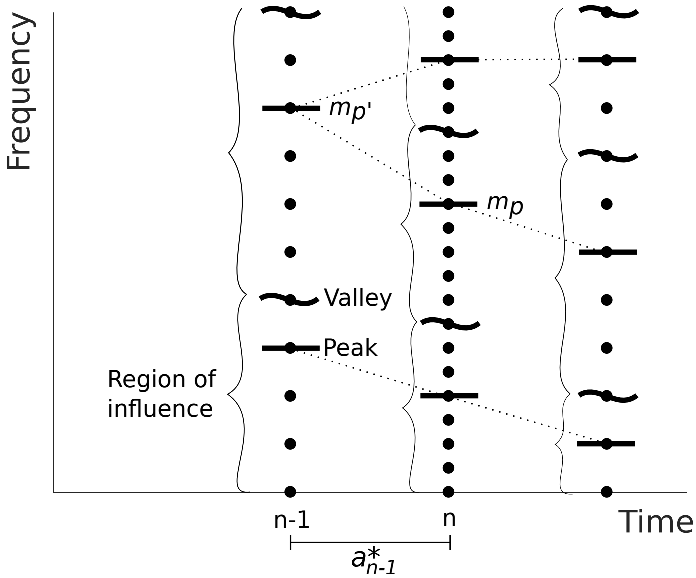

Fig. D.4: Illustration of peak, valley and region of influence.

Based on this construction, with a^*_{n-1} given in (D.9), the phase estimate at peak channel m_p is

$$d\{n\}(m_p) = \left|c\{n\}(m_p)\right| e^{i(\angle d\{n-1\}(m_{p'}) + \omega a^*_{n-1})}. \tag{D.11}$$

For the neighboring channels in the corresponding region of influence, the phase values will be locked to the phase of the peak. Following the approach in [17], we let

$$e^{i\angle d\{n\}(m)} = e^{i\left(\angle d\{n\}(m_p)+\angle c\{n\}(m)-\angle c\{n\}(m_p)\right)},$$

for all channels m in the region of influence corresponding to peak channel m_p. Hence, the phase locking is such that the difference in synthesis phase is the same as the difference in analysis phase. It is important to note that the actual phase estimates are done only at peak channels, which allows for a fast implementation. As mentioned in Section 3.4, the phase estimate (D.11) is not well suited for modelling attack transients. In the next section we explain our approach for dealing with this issue.

Transient preservation

Since the phase values $\angle d\{n\}(m)$ at transients locations cannot be predicted from previous estimates, one might choose to simply reinitialize all phase values at such locations $\angle d\{n\}(m) = \angle c\{n\}(m)$. However, for stationary partials passing through the transient, such a reinitialization completely destroys the horizontal phase coherence, thereby producing undesirable artifacts in the resulting sound. To deal with this problem, we propose the following rule for phase estimation at transient locations: Assume time-instant n corresponds to the onset of an attack transient. Consider channel m, belonging to the region of influence dominated by a peak channel m_p, and let $m_{p'}$ denote the peak channel of the region of influence that channel m_p would have belonged to in time frame $n-1$ (same notation as in (D.11), see also Figure D.4). Then, given a tolerance $\varepsilon > 0$, we reinitialize $\angle d\{n\}(m) = \angle c\{n\}(m)$ if and only if

$$|c\{n\}(m)| > \left|c\{n-1\}(m_{p'})\right| + \varepsilon. \tag{D.12}$$

For the implementation, the calculations are done on a dB scale with $\varepsilon = 2$dB. We note that in contrast to previously proposed techniques for onset reinitialization [3, 7, 26], our algorithm has the advantage that it tracks sinusoids *across* frequency channels.

4.3 Synthesis

Before we can provide the actual synthesis formula, we need to return to the issue of choosing the overlap between window functions (cf. Section 4.1). Originally, scale frames were invented with the intention of construction adaptive TF representations with a very low redundancy. To ensure a low redundancy, and a stable reconstruction, the overlap between adjacent window functions is chosen as 1/3 of the length for equal windows and 2/3 of the length of the shorter window for different windows [1].

This construction makes sense in the general settings, since the resulting system constitutes a frame for $\mathbb{C}^L$ as long as the painless condition from Proposition 2.1 is satisfied. However, in the case of time-stretching with a factor $r > 1$, this construction cannot guarantee that the dual windows (cf. Proposition 2.1) overlap coherently when placed at the synthesis time instants. To tackle this issue, we have chosen the overlap between window functions in the following way:

1. First the onsets of attack transients are calculated (using the onset detection algorithm from Section 4.1).
2. Then these onsets are relocated such that the distance between the relocated onsets is r times the distance between the original onsets.
3. The window functions are now calculated according to the relocated onsets, using the approach in [1], and afterwards centered at the original time instants.

While this approach might give the impression that we just stretch the window lengths by a factor of r, this is not the case. Calculating the windows with respect to the relocated onsets still produce a sequence of windows functions of the same lengths as if the original onsets had been used. This is illustrated in Figure D.5.

With this choice of overlap, we can construct the stretched signal f_* using the synthesis formula

$$f_* = \sum_{n \in \mathbb{Z}_N} \sum_{m \in \mathbb{Z}_{M_n}} d\{n\}(m) \tilde{g}_{m,n}, \tag{D.13}$$

with $\{\tilde{g}_{m,n}\}_{m,n}$ being the canonical dual frame from Proposition 2.1 constructed using the synthesis time instants. In practice, the reconstruction formula (D.13) is realized by applying an inverse FFT and overlap-add as in the classical PV.

4.4 Advances

In this section we explain how the proposed algorithm improves the techniques of the PV. We do so by separately addressing the three drawbacks described in Section 3.4.

1. **Vertical coherence:** If a sinusoid moves from channel $m_{p'}$ at time $n-1$ to channel m_p at time n, then the corresponding peak channel also changes from $m_{p'}$ to m_p. The estimate given in (D.11) therefore ensures that the corresponding phase increment takes this behavior into

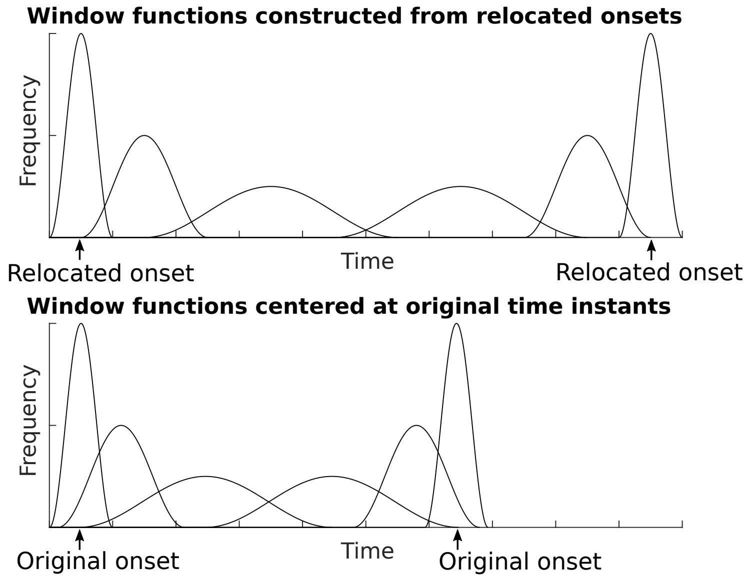

Fig. D.5: Construction of the overlap between window functions.

account. In this way we get coherence *across* the various frequency channels in contrast to the standard PV which only provides coherence *within* each frequency channel.

2. **Resolution:** Changing the representation from that of a DGT to an adaptable NSGT automatically improves the TF resolution for signals, which are well represented by the sinusoidal model (D.5). Furthermore, calculating the phase increment only at peak channels replaces the underlying assumption of well resolved sinusoids in each frequency channel with the weaker assumption of well resolved sinusoids in each region of influence.

3. **Transients:** To reduce transient smearing, we keep the stretch factor equal to one during attack transients and we reinitialize the phase values of relevant channels according to (D.12).

While the PV serves as a good starting point for understanding the fundamental concepts behind the proposed algorithm, it is not the main goal of this article only to improve the resulting sound quality compared to this classical technique. The main advantage of the proposed algorithm is the ability to produce good results, when compared to state of the art, while keeping a low redundancy of the applied TF transform.

Redundancy of the NSGT

As mentioned in Section 3.3, the classical PV applies an overlap of 75% corresponding to a redundancy of 4 in the DGT. There is some mathematical justification to this choice [17], but mainly the overlap is chosen to ensure a good TF resolution. It should be noted that the redundancy of the DGT is independent of the signal under consideration — it only depends on the analysis hop size and the length of the window function (assuming the painless condition is satisfied).

For multi-resolution methods, the situation changes as the TF resolution adapts to the particular signal. A standard approach for multi-resolution methods is to choose non-uniform sampling points in time, with corresponding window functions, and a *uniform* number of frequency channels corresponding to the length of the largest window function [19, 21]. This construction corresponds to applying a uniform NSGF (cf. Section 2). Such an approach is desirable from a practical point of view as the coefficients then form a matrix and the standard techniques from the PV (and its improvements) immediately apply. However, the choice of a uniform NSGF naturally implies a high redundancy of the transform as the sampling density is much higher than needed for the painless case (cf. Proposition 2.1). For real world signals, such a high redundancy is undesirable as it implies a high computational cost for the time-stretching algorithm.

In contrast to previously suggested methods, our algorithm takes full advantage of the painless condition and produces good results with a redundancy of ≈ 3 for a stretch factor of $r = 1.5$. It is important to note that the redundancy of the proposed algorithm depends *both* on the signal under consideration and the stretch factor (at least in the case where $r > 1$). For different signals, the onset detection algorithm calculates different onsets, which results in different time sampling points and different numbers of frequency channels. As for the stretch factor, we recall the choice of overlap as described in Section 4.3. For a large stretch factor, we need a large overlap between the window functions to guarantee that the synthesis formula (D.13) makes sense. We do not consider the dependency between the redundancy and the stretch factor a problem, since the redundancy is still manageable even for large stretch factors. For a stretch factor of $r = 3$, the redundancy is ≈ 5 and for a stretch factor of $r = 4$, the redundancy is ≈ 7.

In the next section we present the numerical experiments and compare the proposed algorithm with state of the art algorithms for time stretching (cf. Section 1.1).

5 Experiments

The proposed algorithm has been implemented in MATLAB R2017A and the corresponding source code is available at

`http://homepage.univie.ac.at/monika.doerfler/NSPV.html`

The source code depends on the following two toolboxes: The LTFAT [23] (version 2.1.2 or above) freely available from `http://ltfat.github.io/` and the NSGToolbox [1] (version 0.1.0 or above) freely available from `http://nsg.sourceforge.net/`.

For the classical PV, we use an implementation by Ellis [11], which includes some improvements to the procedure described in Section 3 (in particular, interpolation of magnitudes). As these improvements result in a significantly improved audio quality, we have chosen this implementation for comparison.

In Section 5.1 we compare the proposed algorithm to the classical PV by stretching synthetic (music) signals and in Section 5.2 we turn to the analysis of real world signals and compare the proposed algorithm with the algorithms from Derrien [7] and Liuni et al. [19].

5.1 Synthetic signals

Analysing synthetic signals has the advantage that the perfect stretched version is available and can be used as ground truth. For this experiment, we construct a large number of synthetic signals and compared the performance of the proposed algorithm with the classical PV for each of these signals. More precisely, the approach is as follows:

1. For each synthetic melody we choose a random number of notes between 4 and 10. Each note has a randomly chosen duration of either 0.5 or 1 second and the corresponding tone consists of a fundamental frequency and three harmonics of decreasing amplitudes. The fundamental frequencies are set to coincide with those of a piano and the melody is allowed to move either 1 or 2 half notes up or down (randomly chosen) per step. A randomly chosen envelope ensures that the tones have both an attack and a release. The sampling frequency of the resulting signal s is 16000 Hz.

2. A stretch factor $0.5 \leq r \leq 3.75$ is chosen at random and another synthetic signal s_{perf} is constructed, such that s_{perf} corresponds to a perfectly time stretched version of s in the sense of (D.6). The classical PV and the proposed algorithm are applied to the original signal s, with respect to the stretch factor r, resulting in the time stretched versions s_{pv} and s_{nsgt}.

3. Three DGTs S_{perf}, S_{pv} and S_{nsgt} are constructed from the time stretched versions s_{perf}, s_{pv} and s_{nsgt}, using the same parameter settings for each signal. With $|S|$ denoting a vector consisting of the absolute values of a DGT S, we use the following error measure

$$E(S_{perf}, S) = \frac{\left\| \left|S_{perf}\right| - |S| \right\|_2}{\left\| \left|S_{perf}\right| \right\|_2}, \tag{D.14}$$

with S being either S_{pv} or S_{nsgt}.

Note that we cannot apply a sample by sample error measure in the time domain, since in this case a small change in phase for the stretched signals might cause a large error, which does not reflect the actual sound quality. We therefore choose to compare the stretched versions using the magnitude difference of their DGTs. Let us now define the parameters used for the TF representations in this experiment.

Choice of parameters

For the DGT used in the PV, we apply two different parameter settings. Using the notation (hopsize, number of frequency channels) we use the parameters $(256, 1024)$ and $(128, 512)$. For the first parameter setting we use a Hanning window of length 1024 and for the second parameter setting we use a Hanning window of length 512. In this way we obtain painless DGTs of redundancy 4.

For the NSGT used in the proposed algorithm, we use 5 different Hanning windows with lengths varying from 96 samples (at attack transients) to $96 \cdot 2^4 = 1536$ samples. The lower bound on the number of frequency channels is set to $96 \cdot 2^3 = 768$, corresponding to the length of the second largest window functions.

For the DGT used for computing S_{perf}, S_{pv}, and S_{nsgt}, we use parameters $(128, 2048)$ and a Hanning window of 2048 samples. This results in a painless DGT of redundancy 16.

Results

Repeating the experiment described above for 1000 synthetic test signals we get the average results shown in Table D.1.

The results in Table D.1 show that the proposed algorithm outperforms the classical PV, with respect to the error measure in (D.14), while keeping a comparable redundancy of the applied transform. For a visualization of the performances of the algorithms we have plotted, in Figure D.6, the spectrograms corresponding to the three DGTs S_{perf}, S_{pv} (with parameters $(128, 512)$), and S_{nsgt} for one particular synthetic test signal (with $r = 1.5$).

Table D.1: Average results for 1000 synthetic test signals.

Algorithm:	PV(256, 1024)	PV(128, 512)	Proposed algorithm
Average red.:	3.954813	3.977300	3.637370
Average error:	0.439982	0.415139	0.095104

We can easily see how the proposed algorithm more accurately reproduces the onsets, and how it reduces the noisy components between the harmonics compared to the PV. However, we can also see how the frequencies corresponding to the harmonics are better reproduced with the PV than with the proposed algorithm. The proposed algorithm induces a certain amplitude modulation due to the peak detection and phase locking approach described in Section 4.2.

We have provided sound files on-line for the test signal shown in Figure D.6 with respect to the stretch factors $r = [0.75, 1.25, 1.5, 2.25, 3.0, 3.75]$. The sound files are available for the perfect stretched version, the PV(128, 512), and the proposed algorithm. It is important to note that the error measure given in (D.14) is not a direct reflection of the actual audio quality — it is for instance not true that the proposed algorithm consistently performs 4 times as good as the classical PV. The results for the proposed algorithm are particularly convincing for stretch factors $r \leq 2$, where the timbre at attack transients is nicely preserved in contrast to the classical PV. However for larger stretch factors $r \geq 2$, the impact of the amplitude modulation, and of the coarse frequency resolution around onsets, becomes audible. Eventually, this results in an overall sound quality comparable to the PV (or even below for very large stretch factors $r \geq 3$).

Since the authors do not have access to the source code of the more sophisticated algorithms as proposed in [3, 7, 19], the comparison for synthetic signals could only be done for the PV and the proposed algorithm. However, as the authors from [7] and [19] kindly provided us with sound files for real world signals, we have included these algorithms for the comparison in the next section.

5.2 Real world signals

For this experiment we consider three real world signals, each of length ≈ 4 seconds and with a sampling frequency of 44100 Hz. The signals are chosen such that they challenge different aspects of the time stretching algorithms:

1. The first signal is a glockenspiel signal with few transients and many harmonics at the higher frequencies.

2. The second signal is a piece of piano music consisting of a dense set of

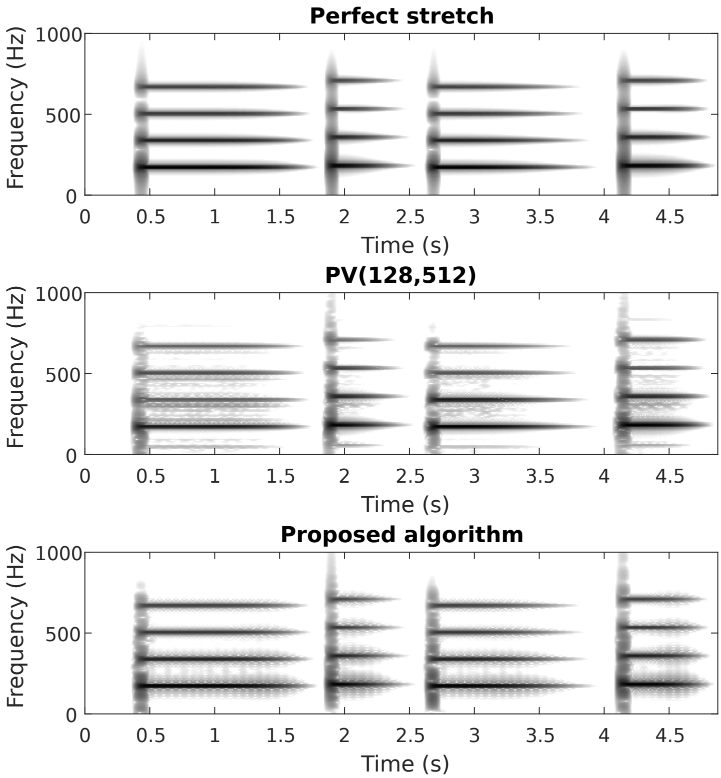

Fig. D.6: Spectrograms for stretched versions of a synthetic signal with $r = 1.5$.

transients with most of the energy concentrated at the lower frequencies.

3. The third signal is from a rock song played by a full band, thereby producing a complex polyphonic sound.

We chose to work with the stretch factors 0.75, 1.25, 1.5 and 2.25 for the comparison. The algorithms we include are:

1. The PV as described in Section 3 and implemented in [11]. For the DGT used in the PV, we use parameters $(512, 2048)$ and a Hanning window of length 2048.

2. The proposed algorithm from Section 4. We use 5 Hanning windows with lengths varying from 384 to $384 \cdot 2^4 = 6144$ and with $384 \cdot 2^2 =$

1536 being the lower bound on the number of frequency channels.

3. The matching pursuit algorithm by Derrien [7].

4. The SuperVP from IRCAM based on the theory of Röbel [26] and Liuni et al. [19]. The algorithm uses only one frequency band and chooses between window lengths of 1024, 2048, 3072, and 4096 samples for the adaptive (uniform) NSGT. We refer the reader to [19] for details.

Since all the stretched sounds are available on-line, we only give the main conclusions. The classical PV and the algorithm by Derrien are rather similar in performance — they both produce a good overall sound quality but with significant transient smearing. The proposed algorithm, on the other hand, does a much better job of preserving the original timbre at attack transient, but induces a certain "roughness" to the sounds (mainly for $r = 2.25$). Also, some of the weaker transients, which are not detected by the onset detection algorithm, suffer from transient smearing for the proposed algorithm (in particular, the "tapping" noises in the background of the piano music). The SuperVP does not have this problem as the transient detection algorithm works on the level of spectral bins. Overall, the SuperVP provides the best audio quality for the three signals, which is to be expected as it applies a TF representation of much higher redundancy than the other algorithms. Calculating the average redundancies for the proposed algorithm (over the 4 stretch factors) for each signal we get 2.40, 2.90 and 2.65. Finally, let us note that the third signal (the rock band signal) reveals a fundamental issue with the application of NSGFs. For $r = 2.25$, neither the proposed algorithm nor the SuperVP are capable of maintaining a steady bass, which results from the changing window lengths. This particular issue is better resolved by the classical PV as well as the algorithm by Derrien.

6 Conclusion and perspectives

Using discrete Gabor theory we have presented the classical PV and proposed a new time stretching algorithm in a unified framework. This approach has allowed us to address and improve on the disadvantages of the classical PV, while preserving the mathematical structure provided by Gabor theory. The proposed algorithm is the first attempt to use non-uniform NSGFs for time-stretching, which allows for a low redundancy of the adaptive TF representation and leads to a fast implementation. The proposed algorithm has been compared to other multi-resolution methods, in a reproducible manner, and we have discussed its advantages and its shortcomings. As a future improvement it could be interesting to connect the techniques presented in this article with the ideas proposed by Röbel in [26], possibly allowing for an algorithm that uses non-uniform NSGFs without the need for fixing the stretch factor during attack transients.

Acknowledgement

We would like to thank the three anonymous reviewers for their suggestions, which clearly improved the over-all presentation of this manuscript. Also a special thanks to Olivier Derrien and Marco Liuni for providing us with the sound files used for comparison.

References

[1] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco. Theory, implementation and applications of nonstationary Gabor frames. *J. Comput. Appl. Math.*, 236(6):1481–1496, 2011.

[2] G. T. Beauregard, M. Harish, and L. Wyse. Single pass spectrogram inversion. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pages 427–431, 2015.

[3] J. Bonada. Automatic technique in frequency domain for near-lossless time-scale modification of audio. In *Proceedings of International Computer Music Conference*, pages 396–399, 2000.

[4] F. Charpentier and M. Stella. Diphone synthesis using an overlap-add technique for speech waveforms concatenation. In *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 2015–2018, Apr 1986.

[5] O. Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], second edition, 2016.

[6] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5):1271–1283, 1986.

[7] O. Derrien. Time-scaling of audio signals with multi-scale Gabor analysis. In *DAFx'07*, pages CD–ROM (6 pages), Bordeaux, France, Sept. 2007.

[8] S. Dixon. Onset detection revisited. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 133–137, Montreal, Quebec, Canada, sep 2006.

[9] M. Dörfler and E. Matusiak. Nonstationary Gabor frames—existence and construction. *Int. J. Wavelets Multiresolut. Inf. Process.*, 12(3):1450032, 18, 2014.

[10] D. Dorran and R. Lawlor. An efficient phasiness reduction technique for moderate audio time-scale modification. In *in Proc. of the 7 th International Conference on Digital Audio Effects (DAFx-04)*, 2004.

[11] D. P. W. Ellis. A phase vocoder in Matlab, 2002. Web resource.

[12] J. L. Flanagan and R. M. Golden. Phase vocoder. *The Bell System Technical Journal*, 45(9):1493–1509, Nov 1966.

[13] D. Griffin and J. Lim. Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):236–243, Apr 1984.

[14] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[15] H. Ishizaki, K. Hoashi, and Y. Takishima. Full-automatic dj mixing system with optimal tempo adjustment based on measurement function of user discomfort. In *ISMIR*, pages 135–140. International Society for Music Information Retrieval, 2009.

[16] J. Laroche. *Time and Pitch Scale Modification of Audio Signals*, pages 279–309. Springer US, Boston, MA, 2002.

[17] J. Laroche and M. Dolson. Improved phase vocoder time-scale modification of audio. *IEEE Transactions on Speech and Audio Processing*, 7(3):323–332, May 1999.

[18] M. Liuni and A. Röbel. Phase vocoder and beyond. *Musica/Tecnologia*, 7:73–89, 2013.

[19] M. Liuni, A. Röbel, E. Matusiak, M. Romito, and X. Rodet. Automatic adaptation of the time-frequency resolution for sound analysis and re-synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5):959–970, May 2013.

[20] R. McAulay and T. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(4):744–754, Aug 1986.

[21] E. Moulines and J. Laroche. Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Commun.*, 16(2):175–205, feb 1995.

[22] M. Portnoff. Implementation of the digital phase vocoder using the fast Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(3):243–248, Jun 1976.

[23] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs. The Large Time-Frequency Analysis Toolbox 2.0. In M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, editors, *Sound, Music, and Motion*, Lecture Notes in Computer Science, pages 419–442. Springer International Publishing, 2014.

[24] M. Puckette. Phase-locked vocoder. In *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, pages 222–225, Oct 1995.

[25] J.-C. Risset. Examples of the musical use of digital audio effects. *Journal of New Music Research*, 31(2):93–97, 2002.

[26] A. Röbel. A new approach to transient processing in the phase vocoder. In *6th International Conference on Digital Audio Effects (DAFx)*, pages 344–349, London, United Kingdom, Sept. 2003.

[27] S. Roucos and A. Wilgus. High quality time-scale modification for speech. In *ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 10, pages 493–496, Apr 1985.

[28] P. L. Søndergaard. *Finite discrete Gabor analysis*. PhD thesis, Technical University of Denmark, 2007.

[29] T. Strohmer. Numerical algorithms for discrete Gabor expansions. In *Gabor analysis and algorithms*, Appl. Numer. Harmon. Anal., pages 267–294. Birkhäuser Boston, Boston, MA, 1998.

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-153-8

AALBORG **UNIVERSITY PRESS**